GABLER EDITION WISSENSCHAFT

Frieder Knüpling

Regimewechselmodelle

Mit einer empirischen Untersuchung von Wechselkursen im Europäischen Währungssystem

DUV

Knüpling
Regimewechselmodelle

GABLER EDITION WISSENSCHAFT

Frieder Knüpling

Regimewechselmodelle

Mit einer empirischen Untersuchung von Wechselkursen im Europäischen Währungssystem

Mit einem Geleitwort
von Prof. Dr. Dietrich Lüdeke

Springer Fachmedien Wiesbaden GmbH

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Knüpling, Frieder:
Regimewechselmodelle : mit einer empirischen Untersuchung von Wechselkursen im Europäischen Währungssystem / Frieder Knüpling. Mit einem Geleitw. von Dietrich Lüdeke.
(Gabler Edition Wissenschaft)
Zugl.: Freiburg (Brsg.), Univ., Diss., 1999
ISBN 978-3-8244-7039-6 ISBN 978-3-663-08918-6 (eBook)
DOI 10.1007/978-3-663-08918-6

Alle Rechte vorbehalten

© Springer Fachmedien Wiesbaden 1999
Ursprünglich erschienen bei Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH, Wiesbaden, und Deutscher Universitäts-Verlag GmbH, Wiesbaden, 1999

Lektorat: Ute Wrasmann / Jutta Hinrichsen

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

http://www.gabler.de
http://www.duv.de

Höchste inhaltliche und technische Qualität unserer Werke ist unser Ziel. Bei der Produktion und Verbreitung unserer Werke wollen wir die Umwelt schonen. Dieses Buch ist deshalb auf säurefreiem und chlorfrei gebleichtem Papier gedruckt. Die Einschweißfolie besteht aus Polyäthylen und damit aus organischen Grundstoffen, die weder bei der Herstellung noch bei der Verbrennung Schadstoffe freisetzen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

ISBN 978-3-8244-7039-6

Geleitwort

Viele, wenn nicht sogar die meisten Wirtschaftsgeschehen, die man mit ökonometrischen Modellen zu erklären versucht, sind im Zeitverlauf durch diskontinuierliche oder kontinuierliche strukturelle Veränderungen gekennzeichnet. In diesen Fällen stellen Modelle mit konstanten, d.h. zeitunabhängigen Parametern insofern keine adäquaten Erklärungsansätze dar, als strukturelle Veränderungen, seien es Veränderungen in den Verhaltensweisen der Wirtschaftssubjekte oder seien es Veränderungen in den Produktionsprozessen, zu Änderungen der Modellparameter führen.

Seit den sechziger Jahren sind zahlreiche Modellansätze entwickelt worden, die solche Veränderungen über geeignete Modellierungen zu erfassen versuchen. Darunter sind insbesondere solche Ansätze von Interesse, bei denen die Parametervariation systematisch ist und einem einfachen Prinzip folgt. Zum einen sind solche Modelle, deren Parametervariation von einem hohen Grad an Regelmäßigkeit gekennzeichnet ist, den klassischen ökonometrischen Modellen nahe verwandt und erlauben den Rückgriff auf die entsprechenden Schätz-, Test- und Prognoseverfahren. Zum anderen läßt sich das Prinzip der Veränderung der Parameter hier häufig inhaltlich interpretieren und gestattet die Konfrontation der Modelle mit wirtschaftstheoretischen Hypothesen.

Unter solchen Modellen mit systematischer Parametervariation stellen Regimewechselmodelle eine wichtige Gruppe dar. Bei diesen Modellen unterliegen die Parameter von Zeit zu Zeit abrupten Sprüngen, wobei aber regelmäßig wieder frühere Parameterwerte angenommen werden. Diesen Modelle, deren Parametervariation verschiedenen Gesetzmäßigkeiten folgen kann, ist seit Beginn der neunziger Jahre neue Aufmerksamkeit zuteil geworden, als die entsprechenden Ansätze von Regressions- auf Zeitreihenmodelle übertragen wurden.

In der vorliegenden Arbeit werden die verschiedenen Erweiterungen der Theorie zum ersten Mal wieder systematisch in einem einheitlichen Rahmen dargestellt. Darüber hinaus werden neue methodische Ansätze, insbesondere für die bisher vernachlässigte Testtheorie, präsentiert. Die vorgestellten Methoden werden an verschieden Anwendungsbeispielen zur Problematik der Modellierung von Wechselkursen im EWS exemplarisch durchgeführt. Diese empirischen Anwendungen sind für den Leser, der sich anhand dieser Arbeit mit den neueren Methoden für Regimewechselmodelle vertraut machen möchte, eine wichtige didaktische Hilfe. Sie sind aber auch von eigenständiger Bedeutung und leisten einen interessanten Beitrag für die Analyse der jüngeren Phase des EWS.

Prof. Dr. Dietrich Lüdeke

Vorwort

Diese Arbeit entstand größtenteils während meiner Zeit als wissenschaftlicher Angestellter an der Abteilung für Statistik und Ökonometrie des Instituts für Allgemeine Wirtschaftsforschung der Albert-Ludwigs-Universität Freiburg. Mein besonderer und herzlicher Dank gilt meinem Betreuer Herrn Prof. Dr. Dietrich Lüdeke für seine verständnisvolle Unterstützung und Förderung. Er war für die offene und freundliche Atmosphäre an der Abteilung verantwortlich, in der diese Arbeit gedeihen konnte. Herrn Prof. Dr. Siegfried Hauser danke ich für die Übernahme des Zweitgutachtens.

Meinen Kollegen Dr. Oliver Schwarz, Wolfgang Hummel, Petra Stellmach und Dr. Bernhard Pfaff danke ich für die unzähligen kleinen und großen Hilfeleistungen, ohne die ich mich aus vielen geistigen und technischen Fußangeln nur schwerlich hätte befreien können. Allen Mitgliedern der Abteilung bin ich zu Dank für ihre Rücksichtnahme und weitgehende Freistellung von universitären Verpflichtungen während der letzten Monate meiner Arbeit an der Dissertation verpflichtet.

Einige der in der Arbeit dargestellten Ergebnisse wurden auf der EC^2-Konferenz über „Simulation Methods in Econometrics“ am 12.-14. Dezember 1996 am „Centro Interuniversitario di Econometria“ (CIDE) und der Universität Florenz präsentiert. Die sich daraus ergebenden Diskussionen und Anregungen waren sehr hilfreich. Von Ian Marsh und Philippe Protin stammen wertvolle technische Hinweise zur Implementierung der Schätzverfahren.

Ich danke der Studienstiftung des Deutschen Volkes, die meine Arbeit zeitweise mit einem Promotionsstipendium unterstützt hat, sowie der Wissenschaftlichen Gesellschaft in Freiburg im Breisgau, die die Veröffentlichung gefördert hat.

Frieder Knüpling

Inhaltsübersicht

1 Einleitung 1

2 Regimewechselmodelle 4

3 Wechselkurse im Europäischen Währungssystem 21

4 ML-Schätzung 35

5 ML-Schätzung von Regimewechselmodellen für Wochendaten des FF/DM-Wechselkurses 69

6 Asymptotische Eigenschaften der ML-Schätzer und Tests 84

7 Tests von Regimewechselmodellen für Wochendaten des FF/DM-Wechselkurses 111

8 Prognosen 117

9 Prognosen für Wochendaten des FF/DM-Wechselkurses 122

10 Weitere empirische Ergebnisse und ökonomische Implikationen 127

11 Schlußbemerkungen 137

Literaturverzeichnis 139

Personen- und Sachregister 148

Inhaltsverzeichnis

1 Einleitung **1**

2 Regimewechselmodelle **4**

2.1 Modelle mit beobachtbaren Zuständen: Schwellenmodelle 6

2.2 Modelle mit latenten Zuständen: Mischungs- und Markov-Modelle 8

2.2.1 Homogene Modelle 8

2.2.2 Inhomogene Modelle 11

2.3 Systematik der Modelltypen 14

2.4 Weitere Modelle mit veränderlichen Parametern 15

2.4.1 Regimewechselmodelle als Modelle mit variablen Parametern 16

2.4.2 Homogene Markov-Modelle als Zustandsraummodelle 17

Anhang: Markov-Ketten 18

3 Wechselkurse im Europäischen Währungssystem **21**

3.1 Das Europäische Währungssystem 21

3.1.1 Geschichtlicher Hintergrund und Entstehung des EWS 21

3.1.2 Die Funktionsweise des EWS 22

3.1.2.1 Der ECU und das Paritätengitter 22

3.1.2.2 Der Interventionsmechanismus 23

3.1.2.3 Das Verfahren zur Neufestlegung der Paritäten 24

3.1.2.4 Die Entwicklung des EWS von 1979 bis zur EWU 24

3.1.2.5 Der Wechselkursmechanismus seit dem 1. Januar 1999 25

3.2 Theoretische Grundlagen für Wechselkursmodelle 26

3.2.1 Determinanten des Wechselkurses 27

3.2.1.1 Kaufkraftparitätentheorie 27

3.2.1.2 Zinsparitätentheorie 28

3.2.2 Monetäre Wechselkursmodelle für flexible Wechselkurse 30

3.2.3 Modelle für Wechselkurse in Bandbreitensystemen 31

3.2.3.1 Modelle vom Krugmann-Typ 31

3.2.3.2 Modelle für den Verlauf von Wechselkursen im EWS 31

3.3 Die empirische Untersuchung von Wechselkursen im EWS mit Hilfe von Regimewechselmodellen 33

4 ML-Schätzung ... 35
4.1 Stochastische Spezifikation ... 35
4.2 Schwellenmodelle ... 38
4.2.1 Bekannter Schwellenwert ... 40
4.2.1.1 Bestimmung der Normalgleichungen ... 40
4.2.1.2 Berechnung von Parameterschätzern durch Hilfsregressionen ... 41
4.2.1.3 Berechnung der ML-Schätzer für heteroskedastische Modelle mit regimeunabhängigen Parametern ... 44
4.2.2 Unbekannter Schwellenwert ... 46
4.3 Mischungs- und Markov-Modelle ... 47
4.3.1 Numerische Maximierung der Likelihood ... 51
4.3.2 Der EM-Algorithmus ... 52
4.3.2.1 Der E-Schritt ... 55
4.3.2.2 Der M-Schritt ... 58
4.3.2.2.1 Die Normalgleichungen für die Regressionsparameter und die Störvarianz(en) ... 58
4.3.2.2.2 Berechnung der neuen Regressionsparameter und Störvarianz(en) durch Hilfsregressionen ... 59
4.3.2.2.3 Berechnung der neuen Regressionsparameter und Störvarianz(en) für heteroskedastische Modelle mit regimeunabhängigen Parametern ... 61
4.3.2.2.4 Berechnung der neuen Parameter des Zustandsprozesses ... 62
Anhang A: Berechnung der geglätteten Wahrscheinlichkeiten der Zustandsvariablen für Markov-Modelle ... 64
Anhang B: Berechnung der neuen Parameterschätzer des Zustandsprozesses für inhomogene Mischungs- und Markov-Modelle ... 66

5 ML-Schätzung von Regimewechselmodellen für Wochendaten des FF/DM-Wechselkurses ... 69
5.1 Ein Schwellenmodell ... 73
5.2 Ein Mischungsmodell ... 75
5.3 Ein Markov-Modell ... 78
Anhang: Verwendete Schätzverfahren ... 81

6 Asymptotische Eigenschaften der ML-Schätzer und Tests ... 84
6.1 Asymptotische Eigenschaften der ML-Schätzer ... 84
6.1.1 Schwellenmodelle ... 84
6.1.1.1 Bekannter Schwellenwert ... 84
6.1.1.2 Unbekannter Schwellenwert ... 90
6.1.2 Mischungs- und Markov-Modelle ... 91
6.2 Tests ... 92
6.2.1 Asymptotische Tests im Kontext von Regimewechselmodellen ... 93
6.2.2 Tests linearer Modelle gegen Regimewechselmodelle ... 95
6.2.2.1 Schwellenmodelle ... 97
6.2.2.1.1 Bekannter Schwellenwert ... 97
6.2.2.1.2 Unbekannter Schwellenwert ... 97
6.2.2.2 Mischungs- und Markov-Modelle ... 98
6.2.3 Tests zwischen Regimewechselmodellen mit verschiedener Dynamik des Zustandsprozesses ... 101
6.2.3.1 Tests zwischen verschiedenen Typen von Schwellenmodellen ... 102
6.2.3.2 Tests zwischen verschiedenen Typen von Mischungs- und Markov-Modellen ... 102
6.2.3.3 Tests von Schwellen- gegen inhomogene Mischungs- und Markov-Modelle ... 103
6.2.4 Spezifikationstests ... 105
6.2.4.1 Schwellenmodelle ... 106
6.2.4.1.1 Bekannter Schwellenwert ... 106
6.2.4.1.2 Unbekannter Schwellenwert ... 107
6.2.4.2 Mischungs- und Markov-Modelle ... 107
6.2.5 Tests zwischen getrennten Hypothesen ... 107

7 Tests von Regimewechselmodellen für Wochendaten des FF/DM-Wechselkurses ... 111
7.1 Tests linearer Modelle gegen Regimewechselmodelle ... 111
7.2 Tests zwischen verschiedenen Typen von Regimewechselmodellen ... 112
7.3 Spezifikationstests ... 114
Anhang: Verwendete Testverfahren ... 115

8 Prognosen **117**

8.1 Schwellenmodelle 118

8.2 Mischungs- und Markov-Modelle 120

9 Prognosen für Wochendaten des FF/DM-Wechselkurses **122**

10 Weitere empirische Ergebnisse und ökonomische Implikationen **127**

10.1 Untersuchung der Wochendaten 127

10.1.1 Zusammenfassung der Ergebnisse für den FF/DM-Kurs und ökonomische Schlußfolgerungen 127

10.1.2 Der NG/DM-Kurs 127

10.1.3 Vergleich der Modelle für die beiden Währungen 133

10.2 Eine Empirische Überprüfung der relativen Kaufkraftparitätentheorie am Beispiel von Monatsdaten des FF/DM- und des NG/DM-Kurses 133

11 Schlußbemerkungen **137**

Literaturverzeichnis **139**

Personen- und Sachregister **148**

Abbildungsverzeichnis

Abbildung 1: Logistische Funktion mit verschiedenen Werten für γ und $c = 0$. 12

Abbildung 2: Ein Beispiel mit Sprungfunktionen als Übergangswahrscheinlichkeiten $p^{11}(z_t)$ (oben) und $p^{22}(z_t)$ (unten) 13

Abbildung 3: Systematik der Regimewechselmodelle. Zahlen in eckigen Klammern geben die Anzahl der freien Parameter des Zustandsprozesses s_t an. 15

Abbildung 4: Wechselkurs zwischen FF und DM, Leitkurs und Interventionsgrenzen..... 69

Abbildung 5: Häufigkeitspolygon und Dichte der geschätztenNormalverteilung von r_t 72

Abbildung 6: Differenz zwischen FF- und DM-Zinsen und FF/DM-Wechselkurs..... 73

Abbildung 7: Abfolge der Zustände im Schwellenmodell (senkrechte Linien entsprechen Zustandswechseln)..... 75

Abbildung 8: Häufigkeitspolygon, Dichte der geschätzten Normal- und Mischungsverteilung von r_t 77

Abbildung 9: Geglättete Wahrscheinlichkeit für Zustand 1 (Mischungsmodell)..... 78

Abbildung 10: Geglättete Wahrscheinlichkeit für Zustand 1 (Markov-Modell)..... 80

Abbildung 11: Vergleich der Übergangswahrscheinlichkeiten des homogenen und des inhomogenen Markov-Modells 113

Abbildung 12: Geglättete und prognostizierte Wahrscheinlichkeiten für den Zustand mit der hohen Volatilität (Prognose 1)..... 123

Abbildung 13: 95 %-Prognoseintervalle des linearen und des Markov-Modells (Prognose 1)..... 124

Abbildung 14: Geglättete und prognostizierte Wahrscheinlichkeiten für den Zustand mit der hohen Volatilität (Prognose 2)..... 125

Abbildung 15: 95 %-Prognoseintervalle des linearen und des Markov-Modells (Prognose 2)..... 126

Abbildung 16: Wechselkurs zwischen Gulden und DM, Leitkurs und Interventionsgrenzen..... 128

Abbildung 17: Geglättete Wahrscheinlichkeit für Zustand 1 (Markov-Modell mit AR(1)-Mechanismus in Zustand 1)..... 132

Abbildung 18: Geglättete Wahrscheinlichkeit für Zustand 1 (Markov-Modell)..... 136

Tabellenverzeichnis

Tabelle 1: Statistische Daten der Verteilung von r_t 71

Tabelle 2: Geschätzte Parameter eines Schwellenmodells für r_t 74

Tabelle 3: Geschätzte Parameter eines Mischungsmodells für r_t 76

Tabelle 4: Momente der geschätzten Mischungsverteilung 76

Tabelle 5: Geschätzte Parameter eines Markov-Modells für r_t 79

Tabelle 6: Kritische Werte des LQ-Tests für homogene Mischungsmodelle mit s unabhängigen Restriktionen von identifizierten Parametern mit Hilfe der Approximation von Davies [1987] 100

Tabelle 7: Kritische Werte der LQ-Statistik aus 10.000 Realisationen einer Simulation unter der Nullhypothese eines Schwellenmodells mit bekanntem und der Alternative eines Schwellenmodells mit unbekanntem Schwellenwert 102

Tabelle 8: Kritische Werte des LQ-Tests aus 1.000 Realisationen einer Simulation unter der Nullhypothese eines Schwellenmodells und der Alternative eines Mischungs- und eines Markov-Modells 105

Tabelle 9: Entscheidungsregel des Tests zwischen getrennten Hypothesen 108

Tabelle 10: LQ-Teststatistiken für die Tests des linearen Modells gegen Regimewechselmodelle 111

Tabelle 11: Schätzwerte eines inhomogenen Markov-Modells 113

Tabelle 12: Geschätzte Parameter eines Markov-Modells mit zustandsunabhängigem Mittelwert 115

Tabelle 13: Geschätzte Parameter eines Markov-Modells mit Schätzzeitraum 23. Januar 1987 bis 15. September 1995 / 26. Juni 1998 122

Tabelle 14: Statistische Daten der Verteilung von r_t 129

Tabelle 15: Geschätzte Parameter eines AR(1)-Modells für r_t 130

Tabelle 16: Schätzwerte eines AR(1)-Markov-Modells für den NG/DM-Kurs 130

Tabelle 17: Schätzwerte eines AR(1)-Markov-Modells für den NG/DM-Kurs mit der Restriktion $\varphi_2 = 0$ 131

Tabelle 18: OLS-Schätzergebnisse der relativen PPP (ohne Konstante) 134

Tabelle 19: Schätzergebnisse des Markov-Modells (PPP) für den FF/DM-Kurs 135

Abkürzungsverzeichnis

AR	Autoregressiv
ARCH	Autoregressive Conditional Heteroscedasticity
BHHH	Berndt-Hall-Hall-Hausman
BIP	Bruttoinlandsprodukt
BLUE	Best Linear Unbiased Estimator
BSP	Bruttosozialprodukt
CAPM	Capital Asset Pricing Model
CIP	Covered Interest Parity
CPI	Consumer Price Index
DF	Dickey-Fuller
DFP	Davidson-Fletcher-Powell
DW	Durbin-Watson
ECU	European Currency Unit
EM	Expectation Maximisation
EWS	Europäisches Währungssystem
EWU	Europäische Währungsunion
EZB	Europäische Zentralbank
FF	Französischer Franc
FGLS	Feasible Generalised Least Squares
GARCH	Generalised Autoregressive Conditional Heteroscedasticity
GLS	Generalised Least Squares
LB	Ljung-Box
LM	Lagrange-Multiplikator
LQ	Likelihood-Quotient
ML	Maximum Likelihood
MSE	Mean Square Error
NG	Niederländischer Gulden
OLS	Ordinary Least Squares
PPP	Purchasing Power Parity
SETAR	Self Exciting Threshold Autoregressive
STAR	Smooth Transition Autoregressive
TAR	Threshold Autoregressive
UIP	Uncovered Interest Parity
WKM II	Wechselkursmechanismus II
WPI	Wholesale Price Index

Verzeichnis mathematischer Symbole

$\text{argmax}_\alpha(f(\alpha))$	das α, das die Funktion $f(\alpha)$ maximiert
D_c	Sprungfunktion mit Sprung bei c
E	Erwartungswert
$E(y\|x)$	bedingter Erwartungswert von y gegeben x
E_γ	Erwartungswert, berechnet mit dem Parameter γ
$f(y\|x)$	Wahrscheinlichkeitsdichte von y gegeben x
F_{α,T_1,T_2}	kritischer Wert der F-Verteilung mit (T_1, T_2) Freiheitsgraden zum Niveau α
$\mathbf{I}$	Einheitsmatrix
$I(\eta)$	Informationsmatrix, berechnet mit dem Parameter η
lim	Limes
log	natürlicher Logarithmus
max	Maximum
min	Mimimum
$N(\mu,\Omega)$	Normalverteilung mit Mittelwertvektor μ und Varianz-Kovarianz-Matrix Ω
plim	stochastischer Limes
$P(y\|x)$	Wahrscheinlichkeit von y gegeben x
R^2	Bestimmtheitsmaß
Var	Varianz
$\chi^2(n)_\alpha$	kritischer Wert der χ^2-Verteilung mit n Freiheitsgraden zum Niveau α
Γ	Gammafunktion
Δ	Differenzenoperator
$\xrightarrow{d}$	Konvergenz in Verteilung

Verzeichnis der Variablen und Parameter

c	Schwellenwert eines Schwellenmodells
$\mathbf{I}_t$	zum Zeipunkt t verfügbare Information
m	maximale Laglänge eines autoregressiven Modells
p	$P(s_t = 1)$ (Mischungsmodell)
p^{ij}	$P(s_t = i \mid s_{t-1} = j)$ (Markov-Modell)
$\mathbf{s} = (s_1,\ldots,s_T)'$	Zustandsvariable
S_t	$(s_t, s_{t-1},\ldots,s_1)$
$\mathbf{u} = (u_1,\ldots,u_T)'$	Störprozeß
$\mathbf{X}$, $\mathbf{W}$	Matrix der exogenen (vorherbestimmten) Variablen (Regressoren)
$\mathbf{X}_t$	t-te Zeile von $\mathbf{X}$
$\mathbf{X}^1$	Matrix, deren t-te Zeile gleich $\mathbf{X}_t$ ist, wenn $s_t = 1$, und $\mathbf{0}$ sonst
$\mathbf{y} = (y_1,\ldots,y_T)'$	endogene (abhängige) Variablen
$\hat{y}_{T+r}$	Prognosewert für y_{T+r}
Y_t	$(y_t, y_{t-1},\ldots,y_1, y_0,\ldots,y_{-m+1})$
$\mathbf{z} = (z_1,\ldots,z_T)'$	Zustandsindikator
α, β_1, β_2	Regressionsparameter
φ_1, φ_2	autoregressive Parameter
ρ	$P(s_1 = 1)$ (Markov-Modell)
σ^2	Fehlervarianz
Ω	Varianz-Kovarianz-Matrix
$\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \ldots$	geschätzte Parameter

Bezeichnung ökonomischer Variablen

$f_{t,t+1}$	Terminkurs zum Zeitpunkt t für den Zeitpunkt $t + 1$
i_t	zeitstetig angepaßter nomineller Zinssatz zum Zeitpunkt t
P_t	Preisniveau zum Zeitpunkt t
r_t	$100(\log w_t - \log w_{t-1}) = 100(\log w_t / w_{t-1})$
w_t	Wechselkurs zum Zeitpunkt t
$w^e_{t,t+1}$	zum Zeitpunkt t für den Zeitpunkt $t + 1$ erwarteter Wechselkurs
w^r_t	realer Wechselkurs zum Zeitpunkt t

1 Einleitung

In verschiedenen ökonomischen Situationen beobachtet man periodisch auftretende, sprunghafte Verhaltensänderungen der beteiligten Subjekte, die den Charakter von diskreten Strukturbrüchen haben. In manchen Fällen tritt eine wiederholte Rückkehr zu früher gezeigtem Verhalten auf, als ob es für die betrachtete Situation verschiedene *Zustände* oder *Regime* (*states, regimes*) des Verhaltens gäbe, wobei zwischen zwei Zustandswechseln stabile Beziehungen zwischen den ökonomischen Variablen beobachtet werden. Beispiele für solches Verhalten sind

- periodische gesamtwirtschaftliche Verhaltensänderungen je nach der momentanen Phase des Konjunkturzyklusses,
- die abwechselnde Preisbestimmung durch Käufer und Verkäufer auf Märkten im Ungleichgewicht ("Käufer- und Verkäufermärkten"),
- das periodische Schließen und Aufgeben von Kartellen, das sich in Preissprüngen äußert,
- spekulatives und nicht-spekulatives Verhalten an Finanzmärkten.

Möchte man in einer derartigen Situation eine oder mehrere Variablen durch Regressions- oder Zeitreihenmodelle erklären, so muß man davon ausgehen, daß die Parameter von dem jeweiligen Zustand abhängen. Solche verallgemeinerten Regressionsmodelle werden in der ökonometrischen Literatur als *Regimewechselmodelle* (*switching regime* oder *switching regression models*) bezeichnet. Seit Beginn der 70er Jahre kommen Regimewechselmodelle in der ökonometrischen Analyse zur Anwendung. Sie stehen wieder verstärkt im Zentrum methodischer und anwendungsorientierter Arbeiten, seit in den 90er Jahren Zeitreihenmodelle mit in dieses Konzept einbezogen wurden.

Es gibt verschiedene Typen von Regimewechselmodellen, die sich in den Gesetzmäßigkeiten, welche für den Wechsel zwischen den Zuständen gelten, unterscheiden, und damit in dem ökonomischen Wirkungsmechanismus, der mit solch einem Modell beschrieben werden soll. Grundsätzlich sind zu unterscheiden

- Modelle mit beobachtbaren Zuständen und
- Modelle mit unbeobachtbaren (latenten) Zuständen.

Im methodischen Teil der Arbeit werden die für die praktische Arbeit wichtigsten Aspekte von Regimewechselmodellen erläutert. Dabei werden einige eher theoretische Punkte, die für Anwendungen von untergeordnetem Interesse sind, bewußt von der Untersuchung ausgeklammert. Dies betrifft zum Beispiel den Zusammenhang von Regimewechselmodellen und dynamischen Systemen oder Fragen der Reversibilität und Invertierbarkeit von Zeitreihen-

Regimewechselmodellen.[1] In diesem Sinne soll die Arbeit einen „*users' guide*" für die Arbeit mit Regimewechselmodellen darstellen.

Weitgehendes Einvernehmen besteht mittlerweile über die schätztechnischen Fragen im Zusammenhang mit diesen Modellen: Obwohl nicht immer einfach zu berechnen, wird mittlerweile fast ausschließlich mit Maximum-Likelihood-Schätzern gearbeitet. Schwierigkeiten bereitet jedoch nach wie vor die Modellauswahl. Selten wird die Wahl eines bestimmten Regimewechselmodells theoretisch motiviert, noch seltener versucht, die Relevanz des gewählten Modells mit statistischen Methoden zu rechtfertigen. Der Hauptgrund hierfür liegt in der mangelnden Entwicklung geeigneter Testverfahren. René Garcia stellte hierzu fest:

> *"While estimation methods for these models are by now well established (...), such is not the case for the testing procedures."*[2]

Diese Lücke soll in dieser Arbeit wenigstens teilweise geschlossen werden. Ein zentrales Anliegen ist dabei, die verschiedenen Typen von Regimewechselmodellen, die häufig je nach Interesse eines Autors einseitig bevorzugt werden, in einen einheitlichen Rahmen zu stellen und dem Anwender objektive Kriterien bei der Auswahl aus einer möglichst breiten Palette von Modellen zur Verfügung zu stellen.

Die hierfür verwendeten Tests beruhen auf den asymptotischen Verteilungen der zugehörigen Maximum-Likelihood-Schätzer, auf die deswegen soweit eingegangen wird, wie es in diesem Rahmen notwendig ist. Abgerundet wird der methodische Teil durch die Behandlung von Prognoseverfahren mit Regimewechselmodellen.

Die Analyse mit Hilfe der verschiedenen Typen von Regimewechselmodellen wird an der Modellierung von Wechselkursdaten aus dem EWS veranschaulicht. Für den Kurs zwischen dem Französischem Franc bzw. dem Niederländischen Gulden und der DM wird die Hypothese von Verhaltensveränderungen, wie sie mit Regimewechselmodellen beschrieben werden können, im Zusammenhang von einfachen Zeitreihenmodellen und der Kaufkraftparitätentheorie untersucht.

Um die theoretischen Ausführungen möglichst direkt mit empirischen Untersuchungen zu veranschaulichen, wurde die Arbeit folgendermaßen strukturiert: Der methodische Teil wurde unterteilt in einen allgemeinen, einführenden Abschnitt über Regimewechselmodelle sowie je einen über die zugehörigen Schätzverfahren, Tests (zusammen mit Ausführungen über die (asymptotische) Verteilung der Schätzer) und Prognoseverfahren. Diese Themenkomplexe werden in den Kapiteln 2, 4, 6 und 8 behandelt.

---

[1] Vgl. hierzu z.B. TONG [1990].

[2] GARCIA [1998], S. 763.

Nach Kapitel 2, der Einführung in die Thematik von Regimewechselmodellen und einer Beschreibung von verschiedenen Anwendungsmöglichkeiten dieser Modelle, folgt in Kapitel 3 eine Einführung in den empirischen Teil der Arbeit, in der die Konstruktionselemente des Europäischen Währungssystems (EWS) und die theoretischen Grundlagen von Wechselkursmodellen für Währungen im EWS dargestellt sowie die Einsatzmöglichkeiten der zuvor behandelten Regimewechselmodelle für solche Daten erläutert werden.

Dieses Wechselspiel zwischen methodischen und empirischen Ausführungen wird in den folgenden Kapiteln der Arbeit fortgesetzt: An jedes der methodischen Kapitel schließt sich eines an, in dem die zuvor erläuterten Verfahren an einem der Anwendungsbeispiele, der Modellierung von wöchentlichen logarithmischen Veränderungsraten des Kurses zwischen dem Französischem Franc und der DM, erläutert werden. Dementsprechend behandeln Kapitel 5, 7 und 9 die Schätzung von Regimewechselmodellen für Wochendaten des Wechselkurses zwischen dem Französischem Franc und der DM, Tests dieser Modelle und Prognosen mit ihrer Hilfe. In Kapitel 10 werden weitere empirische Ergebnisse präsentiert, die zum einen Wochendaten des Kurses zwischen dem Niederländischen Gulden und der DM und zum anderen eine Überprüfung der Kaufkraftparitätentheorie anhand von Monatsdaten des Kurses zwischen dem Französischem Franc bzw. dem Niederländischen Gulden und der DM betreffen. Diese Gliederung führt dazu, daß der empirische Teil über mehrere Kapitel verteilt ist. Dies wird hier um der gewünschten Verzahnung zwischen Theorie und Praxis willen in Kauf genommen.

Im abschließenden Kapitel 11 werden die wichtigsten methodischen und empirischen Ergebnisse der Arbeit zusammengefaßt und einige weiterführende Fragestellungen im Zusammenhang mit Regimewechselmodellen und Wechselkursdaten angesprochen.

2 Regimewechselmodelle

In diesem Kapitel werden die verschiedenen Arten von Regimewechselmodellen anhand ihrer wichtigsten Charakteristika erläutert. Sie werden mit alternativen Modelltypen verglichen und in verschiedene Konzepte der Modellierung mit variablen Parametern eingeordnet. Anhand von veröffentlichten empirischen Untersuchungen wird das weitreichende Spektrum ökonomischer Anwendungsbeispiele beschrieben.

In einem univariaten Regimewechselmodell soll die Variable y_t durch die Variablen

$$\mathbf{X}_t = (x_{t,1},\ldots,x_{t,K})$$

und

$$\mathbf{W}_t = (w_{t,1},\ldots,w_{t,L})$$

erklärt werden. In der vermuteten Regressionsbeziehung

$$y_t = \mathbf{W}_t\alpha + \mathbf{X}_t\beta + u_t$$

sollen sowohl der Parametervektor β als auch die Varianz des Fehlerprozesses u_t von dem gerade herrschenden Zustand abhängen, nicht jedoch der Parametervektor α.[3]

Der Zustand zum Zeitpunkt t wird durch die *Zustandsvariable* s_t beschrieben, die nur endlich viele (in den meisten Anwendungen zwei) Werte annehmen kann, die üblicherweise mit $s_t = 1,2\ldots$ bezeichnet werden. Ein Modell mit zwei Zuständen lautet also

$$y_t = \begin{cases} \mathbf{W}_t\alpha + \mathbf{X}_t\beta_1 + u_t\,, \text{falls } s_t = 1 \\ \mathbf{W}_t\alpha + \mathbf{X}_t\beta_2 + u_t\,, \text{falls } s_t = 2\,, \end{cases}$$

$$\begin{pmatrix} u_1 \\ \vdots \\ u_T \end{pmatrix} \sim N(\mathbf{0},\Omega),\ \Omega = \begin{pmatrix} \sigma^2_{s_1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{s_T} \end{pmatrix}.$$

In der überwiegenden Zahl der theoretischen und empirischen Arbeiten zu Regimewechselmodellen wird davon ausgegangen, daß *alle* Parameter zustandsabhängig sind.[4] Diese Annahme ist aber in vielen Fällen theoretisch nicht gerechtfertigt und widerspricht dem allgemeinen Grundsatz des sparsamen Modellierens (*parsimonious modelling*). Sie wird vermutlich getroffen, weil die Parameter sich in manchen Fällen so leichter schätzen lassen. In dieser Arbeit wird durchgehend der allgemeine Fall betrachtet, in dem zustandsabhängige und -unabhängige Parameter vorliegen.

---

[3] **X** und **W** dürfen natürlich keine gemeinsamen Variablen enthalten, da sonst die entsprechenden Parameter nicht identifiziert sind.

[4] Eine Ausnahme bildet LÜDEKE [1973], sowie LÜDEKE, HUMMEL und RÜDEL [1989], S. 114 ff., im Kontext variabler Parameter. Zu dem Zusammenhang zwischen Regimewechselmodellen und Modellen mit variablen Parametern vgl. Kapitel 2.4.1.

Zur vollständigen Spezifikation des Modells ist festzulegen, welchen Gesetzmäßigkeiten die Zustandsvariable s_t folgt. Hier sind zwei Extremfälle zu unterscheiden:

- die Zustandsvariable hängt direkt (funktional) von einer (oder mehreren) beobachteten Variablen ab, dem sogenannten *Zustandsindikator*
- die Zustandsvariable ist nicht direkt beobachtbar ("latent") und stellt eine zusätzliche stochastische Komponente dar oder repräsentiert fehlende Einflußgrößen in einem nicht vollständig spezifizierten Modell.

Diese beiden Modelltypen werden in den Kapiteln 2.1 und 2.2 behandelt. Modelle, bei denen die Zustandsvariable nicht beobachtet werden kann, werden „in Reinform" in Kapitel 2.2.1 betrachtet. In Kapitel 2.2.2 wird auf Modelle eingegangen, bei denen unvollständige Informationen über den nicht beobachteten Zustand vorliegen. Diese Modelle können aber, wie später erläutert wird, auch als Verallgemeinerungen der in Kapitel 2.1 behandelten Fälle angesehen werden.

Verallgemeinerungen auf mehr als zwei Zustände sind denkbar. Modelle mit mehr als zwei Zuständen lassen sich aber nur selten theoretisch überzeugend rechtfertigen. Zudem benötigen sie eine mit der Zahl der Zustände rapide wachsende Anzahl von freien Parametern. Schon in einem Modell mit zwei Zuständen ist die Zahl der freien Parameter etwa doppelt so hoch wie in einem gewöhnlichen Modell, weswegen die Verwendung eines Regimewechselmodells immer statistisch gerechtfertigt werden sollte. Modelle mit drei oder mehr Zuständen scheinen jedoch in Anwendungen nur selten signifikant besser zu sein als solche mit einem oder zwei Zuständen.[5] Aus diesen Gründen werden Modelle mit mehr als zwei Zuständen in empirischen Untersuchungen nur selten herangezogen[6] und in dieser Arbeit nicht systematisch untersucht.

Außer den hier betrachteten Einzelgleichungsmodellen sind vereinzelt Mehrgleichungsmodelle (vektorautoregressive Modelle) mit mehreren Zuständen in Betracht gezogen worden. Solche Modelle wurden bisher vorwiegend angewendet, um die wechselseitige Beeinflussung des Konjunkturverlaufs mehrerer Volkswirtschaften zu untersuchen.[7]

---

[5] Es ist mit einigen Schwierigkeiten verbunden, die Anzahl der Zustände mit Hilfe von statistischen Tests zu ermitteln (vgl. Kapitel 6).

[6] Beispiele hierfür stellen GARCIA und PERRON [1996] oder SICHEL [1994] dar.

[7] Beispiele für bivariate (Markov-)Modelle finden sich in PHILLIPS [1991].

2.1 Modelle mit beobachtbaren Zuständen: Schwellenmodelle

Die typischen Modelle mit beobachtbaren Zuständen sind *Schwellenmodelle* (*threshold models*). Bei einem Schwellenmodell ändert sich der Zustand, wenn ein Zustandsindikator z_t einen Schwellenwert über- oder unterschreitet.[8] Für ein Schwellenmodell mit zwei Zuständen und Schwellenwert c gilt also

$$y_t = \begin{cases} \mathbf{W}_t\alpha + \mathbf{X}_t\beta_1 + u_t \text{, falls } z_t < c \\ \mathbf{W}_t\alpha + \mathbf{X}_t\beta_2 + u_t \text{, falls } z_t \geq c. \end{cases}$$

Formuliert man die Modellgleichung folgendermaßen um:

$$y_t = D_c(z_t)(\mathbf{W}_t\alpha + \mathbf{X}_t\beta_1) + (1 - D_c(z_t))(\mathbf{W}_t\alpha + \mathbf{X}_t\beta_2) + u_t \text{ ,}$$

mit der Sprungfunktion

$$D_c(z_t) = \begin{cases} 1 \text{, falls } z_t < c \\ 0 \text{, falls } z_t \geq c \end{cases} \text{,}$$

so erkennt man, daß man Schwellenmodelle als nichtlineare Regressionsmodelle mit unstetiger Regressionsfunktion und Heteroskedastie interpretieren kann.

Typische Beispiele für Schwellenmodelle in ökonomischen Zusammenhängen resultieren aus makroökonomischen Verhaltensgleichungen, bei denen ein Teil der Parameter unterschiedliche Werte je nach der momentanen Konjunkturphase annimmt, die sich danach bestimmt, ob ein geeigneter Konjunkturindikator (etwa die Wachstumsrate des BIP) ober- oder unterhalb eines Schwellenwertes liegt.

In manchen Fällen wird man den Schwellenwert c als bekannt ansehen (z.B. $c = 0$ für Wachstumsraten z_t). Dann kann man aus den Werten von z_t bereits die Abfolge der Zustände ablesen. In anderen Situationen ist c unbekannt und muß aus den Daten geschätzt werden. Durch Vergleich des geschätzten Schwellenwertes mit den Werten von z_t läßt sich dann die vermutete Aufteilung der Daten auf die Zustände angeben.

Einen wichtigen Spezialfall erhält man für $z_t = t$, wenn also die Schwellenvariable gleich dem Zeitindex ist. Dann liegt ein Strukturbruchmodell mit bekanntem (unbekanntem) Zeitpunkt des Strukturbruchs vor, falls c bekannt (unbekannt) ist.

Von GOLDFELD und QUANDT [1972, 1973, 1976] stammen die ersten methodischen Arbeiten über Schwellenmodelle mit exogenen Regressoren. Eine der ersten ökonomischen Anwendungen eines Schwellenmodells ist das Modell des US-amerikanischen Immobilien-

[8] Es können auch mehr als zwei Zustände und entsprechend mehrere Schwellenwerte vorliegen. Der Zustandsindikator z_t kann auch eine Funktion (z.B. eine Linearkombination) mehrerer beobachteter Größen mit unbekannten Parametern sein.

marktes von FAIR und JAFFEE [1972], bei dem z_t die Veränderungsrate eines zugehörigen Preisindexes darstellt. Schwellenmodellen ist seit den Arbeiten von TONG [1983, 1990] und anderen über Zeitreihen-Schwellenmodelle (*Threshold Autoregressive Models*, TAR), bei denen die Variable z_t ein Lag von y_t ist (*Self-Exciting Threshold Autoregressive Models*, SE-TAR), neue Aufmerksamkeit zuteil geworden.

Es sind auch andere Mechanismen denkbar, die bestimmen, wie ein Indikator z_t Zustandswechsel determiniert. Zum Beispiel kann es sein, daß niedrige Werte von z_t Zustand 1 entsprechen und hohe Werte Zustand 2, daß aber der Schwellenwert c_1, den der Zustandsindikator für einen Wechsel von Regime 1 in 2 *über*schreiten muß, höher ist als der Schwellenwert c_2, den der Zustandsindikator für einen Wechsel von Regime 2 in 1 *unter*schreiten muß. Solch ein *verallgemeinertes Schwellenmodell* wäre z.B. angebracht, wenn die Zustände die allgemeine (optimistische oder pessimistische) Einschätzung der ökonomischen Lage beschreiben, die erst dann umschlägt, wenn ein wirtschaftlicher Indikator (etwa eine Wachstumsrate o.ä.) deutlich anwächst oder fällt, für moderate Werte des Indikators (zwischen den Schwellenwerten) aber unverändert bleibt.

In diesem Fall kann es vorkommen, daß selbst bei bekannten Schwellenwerten c_1 und c_2 die Abfolge der Zustände nicht zweifelsfrei aus den Werten von z_t abgelesen werden kann. Liegt nämlich der erste beobachtete Wert von z_t zwischen den beiden Schwellenwerten, so sind die ersten Zustände unbestimmt. Die Zustände können erst von dem Zeitpunkt an mit Sicherheit angegeben werden, zu dem der Zustandsindikator zum ersten Mal den unteren Schwellenwert unter- oder den oberen Schwellenwert überschreitet. Sind die Schwellenwerte unbekannt, so ergibt sich wie im Falle gewöhnlicher Schwellenmodelle zusätzlich noch die Notwendigkeit, diese aus den Daten zu schätzen.

Schon GOLDFELD und QUANDT [1973] schlugen vor, die in der obigen Formulierung eines Schwellenmodells verwendete Sprungfunktion durch eine stetige Funktion zu ersetzen, etwa die logistische Funktion. Solche Modelle mit „glattem Übergang" erlauben ein Kontinuum von Zuständen zwischen zwei Extremen und haben in der letzten Zeit in der Zeitreihenform (*Smooth Transition Autoregressive Models*, STAR) wieder verstärktes Interesse gefunden.[9]

[9] Zu STAR-Modellen vgl. z.B. GRANGER und TERÄSVIRTA [1993], TERÄSVIRTA [1994] und TERÄSVIRTA, TJØSTHEIM und GRANGER [1994].

2.2 Modelle mit latenten Zuständen: Mischungs- und Markov-Modelle

2.2.1 Homogene Modelle

Hängen die Zustandswechsel selbst nicht direkt von einer beobachtbaren Größe ab, so bildet die Abfolge der Zustände einen latenten stochastischen Prozeß s_t, $t = 1,...,T$, der die Verteilung der abhängigen Variable beeinflußt. Dabei sind verschiedene Varianten für die Dynamik von s_t möglich. In den meisten Anwendungen wird angenommen, daß s_t von den erklärenden Variablen unabhängig ist. Die entsprechenden Modelle werden als *homogene* Modelle bezeichnet.

Im einfachsten Fall sind die s_t zeitlich unabhängig und identisch verteilt (d.h. stationär), bilden also eine *Bernoullifolge*. Diese Modelle, die als *Mischungsmodelle*[10] bezeichnet werden, verwenden im Falle zweier Zustände einen freien Parameter für den Zustandsprozeß, etwa $p = P(s_t = 1)$.[11]

Ein für Anwendungen in der Finanzmarktökonometrie wichtiger Fall ist der, in dem $\mathbf{X}$ (bzw. $\mathbf{W}$) ein konstanter Spaltenvektor ist und $\mathbf{W} = \mathbf{0}$ (bzw. $\mathbf{X} = \mathbf{0}$) gilt. Hier entstammen die Werte von y_t je nach Wert von s_t einer von zwei Normalverteilungen mit unterschiedlichen Mittelwerten und/oder Fehlervarianzen, und die Verteilung von y_t stellt eine Mischung (*i.i.d. mixture distribution*) aus diesen beiden Verteilungen dar.[12]

Solche Mischungsverteilungen können, je nachdem, welche Werte die bis zu 5 freien Parameter annehmen, Schiefe, Platy- oder Leptokurtosis (d.h. positive oder negative Kurtosis[13]) und auch Bimodalität aufweisen. Insbesondere Leptokurtosis ist ein gesichertes Charakteristikum der verschiedensten Arten von kurzfristigen Finanzmarktdaten (genauer gesagt: der logarithmischen Tages- und Wochenrenditen), das zuerst von MANDELBROT [1963] und FAMA [1965] und in der Folgezeit von zahlreichen weiteren Autoren beschrieben und modelliert wurde.[14] Die empirischen Untersuchungen mit Hilfe von Mischungsverteilungen weisen darauf hin, daß sich die beiden Zustände für die Renditen deutlich in der Varianz unterschei-

[10] Gelegentlich findet sich hierfür im Deutschen auch die Bezeichnung *Überlagerungen*.

[11] Die erste Untersuchung dieser Verteilungen stammt von K. PEARSON [1894], der die Parameter von Mischungsverteilungen für verschiedene Größenmaße von Krebsen (u.a. Kopfgrößen und Zahnabstände) mit Hilfe der Methode der Momente schätzte.

[12] Zu Mischungsverteilungen vgl. EVERITT und HAND [1981] und TITTERINGTON, SMITH und MAKOV [1985]. Die entsprechenden Regressions- und Zeitreihenmodelle werden in der Literatur gelegentlich auch als *simple switching models* oder *i.i.d.-switching models* bezeichnet.

[13] In dieser Arbeit wird als Kurtosis das um 3 verminderte standardisierte 4. Moment einer Verteilung bezeichnet, d.h. die Kurtosis einer Normalverteilung ist 0.

den.[15] Die beobachtete Leptokurtosis scheint also dadurch zu entstehen, daß in den untersuchten Stichproben Zeitpunkte mit geringer und mit hoher Volatilität gemischt sind.[16]

Die am häufigsten untersuchte Alternative zu Mischungsverteilungen stellen stabile (Pareto-)Verteilungen[17] dar, die jedoch keine endlichen Varianzen besitzen und bei Aggregation nicht dem Zentralen Grenzwertsatz unterliegen. Dies widerspricht den empirischen Beobachtungen: Zum einen werden die empirischen Varianzen bei wachsender Beobachtungsgröße nicht größer, zum anderen weisen die (durch Aggregation hervorgegangenen) längerfristigen (Monats-, Quartals-)Renditen geringere bis gar keine Leptokurtosis auf und streben in Verteilung gegen Normalverteilungen. In vergleichenden Untersuchungen der Anpassungsgüte dieser Modellklassen schneiden die Mischungsverteilungen gut ab.[18] Allerdings benötigen die hier meistens verwendeten Mischungsverteilungen mit gleichen Mittelwerten (sogenannte *scale mixtures*) mit 4 freien Parametern einen Parameter mehr als Paretoverteilungen.

In vielen Fällen erscheint es nicht sinnvoll, die Zustände als unabhängig voneinander anzunehmen. Der einfachste stochastische Prozeß mit Autokorrelation und zwei (endlich vielen) Zuständen ist eine *homogene Markov-Kette erster Ordnung*, bei der die Verteilung der Variablen nur von dem zuletzt angenommenen Zustand, nicht aber von höheren Lags abhängt.[19] Ist die Markov-Kette s_t *stationär*, so ist ihre Verteilung vollständig bestimmt durch die Übergangswahrscheinlichkeiten

$$p^{ij} = P(s_t = i \mid s_{t-1} = j) \ .$$

Es gilt $p^{12} = 1 - p^{11}$ und $p^{22} = 1 - p^{21}$, es bleiben also zwei freie Parameter für die Modellierung der Zustandsvariablen. Falls man auf die Stationaritätsannahme verzichtet und die Startverteilung aus den Daten schätzt, so kommt noch ein freier Parameter, etwa $\rho_1 = P(s_0 = 1)$,

[14] Mischungsverteilungen in diesem Zusammenhang verwenden z.B. FIELITZ und ROZELLE [1983], PAN, CHAN und FOK [1995] oder HALL [1996].

[15] Verallgemeinerungen auf mehr als zwei Zustände wären denkbar. Die Untersuchungen weisen jedoch darauf hin, daß zusätzliche Zustände den Erklärungsgehalt der Modelle nicht erhöhen, vgl. z.B. KAEHLER und MARNET [1994], S. 209.

[16] Mischungsverteilungen mit gleichen Mittelwerten und unterschiedlichen Varianzen sind *immer* leptokurtisch. Evt. beobachtete Schiefe läßt sich mit unterschiedlichen Mittelwerten der beiden Verteilungen erklären.

[17] Vgl. z.B. und MITTNIK und RACHEV [1993]. - Weitere in diesem Kontext verwendete Verteilungen sind Student-Verteilungen (vgl. z.B. PRAETZ [1972]), verallgemeinerte Student-Verteilungen (vgl. z.B. LYE, MARTIN und TEO [1998]) und hyperbolische Verteilungen (vgl. z.B. EBERLEIN und KELLER [1995] und EBERLEIN, KELLER und PRAUSE [1998]).

[18] Vgl. etwa KON [1984] und BOOTHE und GLASSMAN [1987]. Diese Arbeiten weisen allerdings einige methodische Unzulänglichkeiten auf, vgl. dazu Fußnote 150 auf S. 98.

[19] Ein Überblick über die wichtigsten für diese Arbeit relevanten Eigenschaften von Markov-Ketten wird im Anhang dieses Kapitels gegeben.

dazu. Die oben erwähnten Mischungsmodelle erhält man aus solchen *Markov-Modellen*, wenn $p^{12} = p^{22}$ (und $\rho_1 = p^{11}$) gilt.

Markov-Modelle haben eine lange Tradition in der Ökonometrie und sind für die Modellierung der verschiedensten Daten verwendet worden. Die ersten methodischen und empirischen Arbeiten in der Ökonometrie, die solche Ansätze verwenden, sind die von GOLDFELD und QUANDT [1972, 1973], in denen Markov-Modelle mit exogenen Regressoren betrachtet werden, die den US-amerikanischen Immobilienmarkt beschreiben. Hier stehen die beiden Zustände für einen „Käufermarkt" und einen „Verkäufermarkt". Markov-Zeitreihenmodelle wurden erstmals von HAMILTON [1988, 1989, 1993] systematisch untersucht, der Modelle für BSP-Zeitreihen vorstellte, in denen die beiden Zustände den expansiven und rezessiven Phasen des Konjunkturzyklusses entsprechen. Ein weiteres häufig zitiertes Modell ist das von COSSLETT und LEE [1985] (mit exogenen Regressoren) für das Preissetzungsverhalten des US-amerikanischen Eisenbahnkartells von 1880 bis 1886, in dem die beiden Zustände abgesprochenes und konkurrierendes Verhalten der Kartellmitglieder beschreiben.[20] ENGEL und HAMILTON [1990] modellierten die wechselnden langfristigen Auf- und Abschwünge des Wechselkurses des Dollars gegenüber den Währungen westlicher Industrienationen mit Hilfe von Markov-Zeitreihenmodellen. CECCHETTI, LAM und MARK [1990] untersuchten verschiedene US-amerikanische makroökonomische Variablen,[21] KAEHLER und MARNET [1994] Devisenkursrenditen mit Markov-Modellen ohne Regressoren.

In empirischen Modellen für kurzfristige Kursrenditen unterscheiden sich die beiden Zustände wie bei den Mischungsverteilungen in aller Regel deutlich in der Volatilität, und die geschätzten Wahrscheinlichkeiten für das Verbleiben in den Zuständen (p^{11} und p^{22}) sind relativ hoch,[22] die durchschnittliche Verweildauer in diesen Zuständen also dementsprechend lang. Diese Modelle beschreiben also die empirisch beobachtete Häufung von Perioden mit hoher bzw. niedriger Volatilität (*volatility clustering*) und stehen damit in Konkurrenz zu ARCH-Modellen und ihren Weiterentwicklungen.[23]

Dies ist nur ein kleiner Ausschnitt aus dem Spektrum von empirischen Anwendungen von Mischungs- und Markov-Modellen. In der ökonometrischen Literatur der letzten Jahre ist das

[20] Vgl. hierzu auch LEE und PORTER [1984].

[21] Vgl. hierzu auch FERRI und GREENBERG [1992] und LAHIRI und WANG [1994].

[22] Vgl. z.B. KAEHLER und MARNET [1994], S. 214 ff. Für die dort untersuchten Tages- und Wochenrenditen von vier Wechselkursen sind diese Wahrscheinlichkeiten durchgängig größer als 0,89.

[23] Vgl. z.B. BOLLERSLEV, CHOU und KRONER [1992] oder BOLLERSLEV, ENGLE und NELSON [1994].

Interesse an ihnen unvermindert, und die Zahl der ökonomischen Anwendungen ist mittlerweile kaum noch zu überschauen.

Für manche empirische Anwendung wäre es denkbar, Modelle mit Markov-Ketten höherer Ordnung (also das diskrete Gegenstück zu AR(p)-Modellen) zu verwenden, bei denen die Verteilung der Zustände von den letzten p Zuständen abhängt ($p \geq 1$). Damit könnte man z.B. die plausible Annahme überprüfen, daß die Wahrscheinlichkeit für das Verbleiben auf einem Wachstumspfad höher ist, wenn die letzten p Perioden schon expansiv waren, als wenn nur die letzte Periode Anzeichen eines Aufschwungs zeigt, die Perioden davor aber rezessiv waren. Solche Modelle spielen trotz ihrer interessanten theoretischen Eigenschaften in der Literatur bisher keine nennenswerte Rolle, da sie deutlich mehr freie Parameter benötigen als Modelle mit Markov-Ketten erster Ordnung.[24]

2.2.2 Inhomogene Modelle

Schon GOLDFELD und QUANDT [1973] schlugen Modelle vor, in denen die Aufenthaltswahrscheinlichkeit p bzw. die Übergangswahrscheinlichkeiten p^{ij} von einer beobachteten Größe z_t funktional abhängen. In diesen Modellen sind die Zustände also latent, der Regimeindikator z_t enthält aber eine gewisse Information über den jeweils herrschenden Zustand. Solche Modelle werden als *inhomogene* Mischungs- bzw. Markov-Modelle bezeichnet.[25]

Es sind verschiedene Formen der Abhängigkeit der Aufenthaltswahrscheinlichkeit p bzw. der Übergangswahrscheinlichkeiten p^{ij} von z_t denkbar. In empirischen Anwendungen werden fast ausschließlich monotone Funktionen in Erwägung gezogen, die für kleine Werte von z_t gegen 0 und für große Werte gegen 1 streben (oder umgekehrt).[26] Beispiele für diesen Funktionstyp sind Verteilungsfunktionen unimodaler Wahrscheinlichkeitsdichten, etwa der Standardnormalverteilung, oder die logistische Funktion

$$f(z_t) = \frac{e^{\gamma(z_t - c)}}{1 + e^{\gamma(z_t - c)}} \ .$$

Der Parameter c ist der "Umschlagpunkt", an dem f den stärksten Anstieg aufweist. Die Stärke der Abhängigkeit der Aufenthalts- bzw. der Übergangswahrscheinlichkeiten von z_t beschreibt der Parameter γ. Inhomogene Mischungsmodelle von diesem Typ verwenden dem-

[24] Man kann eine Markov-Kette höherer Ordnung auffassen als eine Markov-Kette erster Ordnung mit passend vergrößertem Zustandsraum. Deswegen sind die in diesen Modellen zusätzlich auftretenden methodischen Schwierigkeiten dieselben wie für Markov-Modelle mit mehr als zwei Zuständen, vgl. z.B. FRIEDMANN [1994].

[25] Vgl. zur Bezeichnung den Anhang zu diesem Kapitel.

[26] Andere denkbare Funktionstypen wären spiegelsymmetrische, die für weit von c entfernte Werte von z_t klein und in der Nähe von c groß sind (oder umgekehrt), etwa $f(x) = 1/(1-(x-c)^2)$.

nach die beiden Parameter c und γ zur Modellierung des Zustandsprozesses s_t. Für inhomogene Markov-Modelle wären z.B. p^{11} und p^{22} logistische Funktionen mit den freien Parametern c_1, c_2, γ_1 und γ_2.

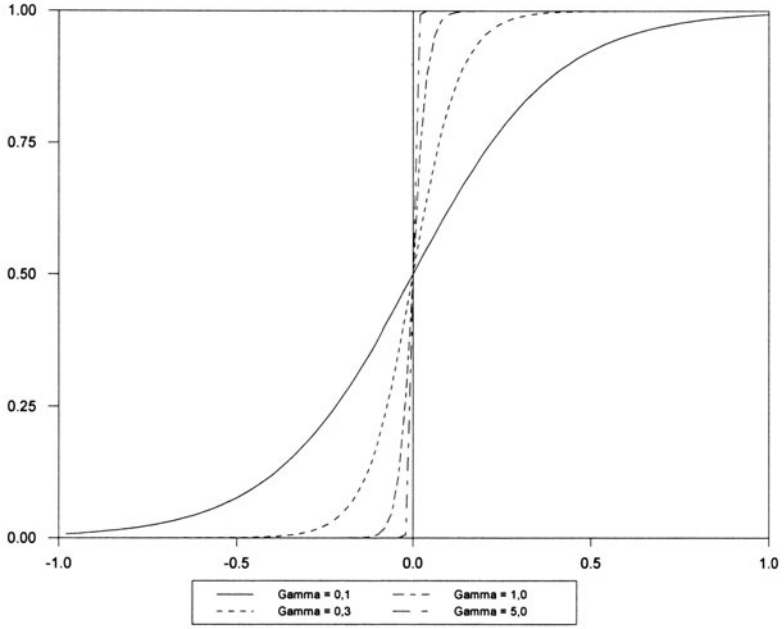

Abbildung 1: Logistische Funktion mit verschiedenen Werten für γ und $c = 0$.

Bei Mischungsmodellen mit solchen Aufenthaltswahrscheinlichkeiten ist also Zustand 1 für (im Vergleich zu c) sehr kleine Werte von z_t nahezu unmöglich und für sehr große Werte von z_t fast sicher. Für Werte in der Nähe von c ist die stochastische Komponente des Modells ausgeprägt. Hier kann der Zustand nicht mit Sicherheit angegeben werden, etwa aufgrund weiterer, nicht im Modell enthaltener Einflußgrößen. Analog lassen sich die Übergangswahrscheinlichkeiten für inhomogene Markov-Modelle interpretieren.

Für sehr große Werte von γ konvergiert die logistische Funktion (von dem Wert c selbst abgesehen) punktweise gegen eine Stufen-(oder Treppen-)Funktion mit einer Unstetigkeitsstelle bei c (vgl. Abbildung 1). Gibt diese logistische Funktion die Aufenthaltswahrscheinlichkeit in Zustand 1 eines inhomogenen Mischungsmodells an, so kann man im Grenzfall den herrschenden Zustand mit Sicherheit angeben – es liegt ein Schwellenmodell mit

Schwellenwert c vor. Schwellenmodelle sind also Spezialfälle inhomogener Mischungs- (und damit auch Markov-)Modelle.

Weitere interessante Spezialfälle inhomogener Markov-Modelle erhält man, wenn man Stufenfunktionen mit den Sprungstellen c_1 und c_2 für die beiden Übergangswahrscheinlichkeiten (etwa p^{11} und p^{22}) wählt. In diesem Fall kann man die Zustands*wechsel* mit Sicherheit angeben. Es sind, je nach Lage der Sprungwerte und der Richtung des Sprunges (von 0 auf 1 oder umgekehrt), verschiedene Konstellationen möglich. Ein interessanter Fall ist der mit dem in Abbildung 2 dargestellten Verlauf der Übergangswahrscheinlichkeiten.

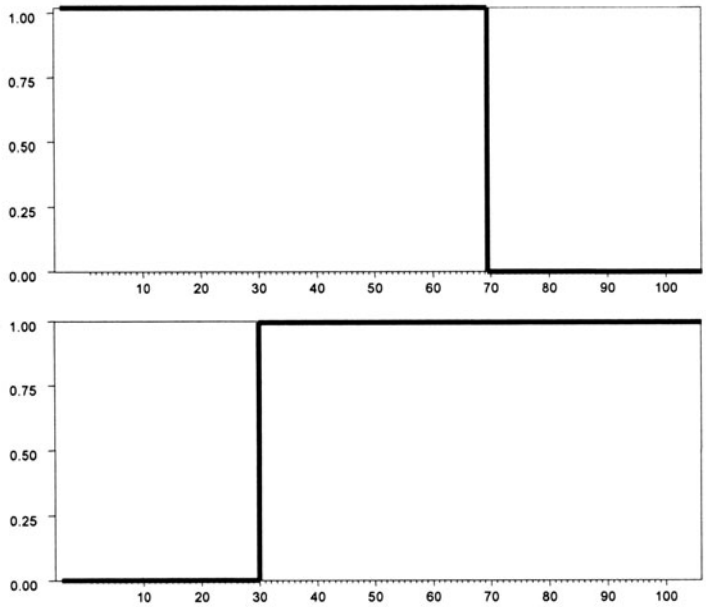

Abbildung 2: Ein Beispiel mit Sprungfunktionen als Übergangswahrscheinlichkeiten $p^{11}(z_t)$ (oben) und $p^{22}(z_t)$ (unten)

Solange der Wert des Indikators zwischen den beiden Schwellenwerten liegt, haben sowohl p^{11} als auch p^{22} den Wert 1, der jeweilige Zustand (sei es nun Zustand 1 oder 2) ändert sich also nicht. Für Indikatorwerte unterhalb des unteren Schwellenwertes bleibt Zustand 1, und Zustand 2 wechselt zu 1; umgekehrtes gilt für Indikatorwerte oberhalb des oberen Schwellenwertes. Es liegt also gerade ein verallgemeinertes Schwellenmodell vor, wie es in Kapitel 2.1 beschrieben wurde.

Auch homogene Markov- und Mischungsmodelle lassen sich natürlich als Spezialfälle aus inhomogenen Modellen zurückgewinnen. Dafür ist eine andere Parametrisierung der logistischen Funktion vorteilhaft, z.B. für Mischungsmodelle:

$$p(z_t) = \frac{e^{\delta+\gamma z_t}}{1+e^{\delta+\gamma z_t}}.$$

Ist hier $\gamma = 0$, so liegt ein homogenes Mischungsmodell mit $p = e^{\delta}/(1+e^{\delta})$ vor. Entsprechendes gilt für Markov-Modelle mit $\gamma_1 = 0$ und $\gamma_2 = 0$.

LEE [1991] und DIEBOLD, LEE und WEINBACH [1994] haben solche inhomogenen Modelle im Gefolge der Renaissance von Markov-Modellen in den 90er Jahren wieder aufgegriffen und wichtige methodische Aspekte dieser Modelle behandelt. In FILARDO [1994] werden Modelle, in denen y den US-amerikanischen Produktionsindex darstellt und z_t verschiedene Konjunkturindikatoren, betrachtet, um so die Prognosekraft dieser Indikatoren zu vergleichen. Hier gilt das Hauptaugenmerk der Tatsache, daß diese Modelle die empirisch beobachteten zeitlich variierenden Verweildauern in den konjunkturellen Phasen (Zuständen) erklären können (im Gegensatz zu gewöhnlichen Markov-Modellen, bei denen im stationären Fall die erwartete Verweildauer eines Zustandes konstant ist).[27] In GHYSELS [1993] wird anhand eines Modells, in dem z_t die Saison (die Quartalsnummer) darstellt, untersucht, ob Wechsel zwischen konjunkturellen Phasen bevorzugt in bestimmten Jahreszeiten stattfinden. In DURLAND und MCCURDY [1994] wird ein Modell vorgeschlagen, in dem z_t die Verweildauer in dem jeweiligen Zustand darstellt. Damit wird untersucht, ob die Wahrscheinlichkeit für das Eintreten einer Rezession mit zunehmender Dauer einer Aufschwungphase zunimmt (und umgekehrt).

2.3 Systematik der Modelltypen

Die Diskussion Kapitel 2.2 hat gezeigt, daß sich alle vorher erwähnten Regimewechselmodelle als Spezialfälle von inhomogenen Markov-Modellen ansehen lassen. Einen Überblick über die wichtigsten Zusammenhänge zwischen den verschiedenen Modelltypen gibt Abbildung 3, in der Parameterrestriktionen durch Pfeile angedeutet werden.

[27] Vgl. hierzu den Anhang zu diesem Kapitel.

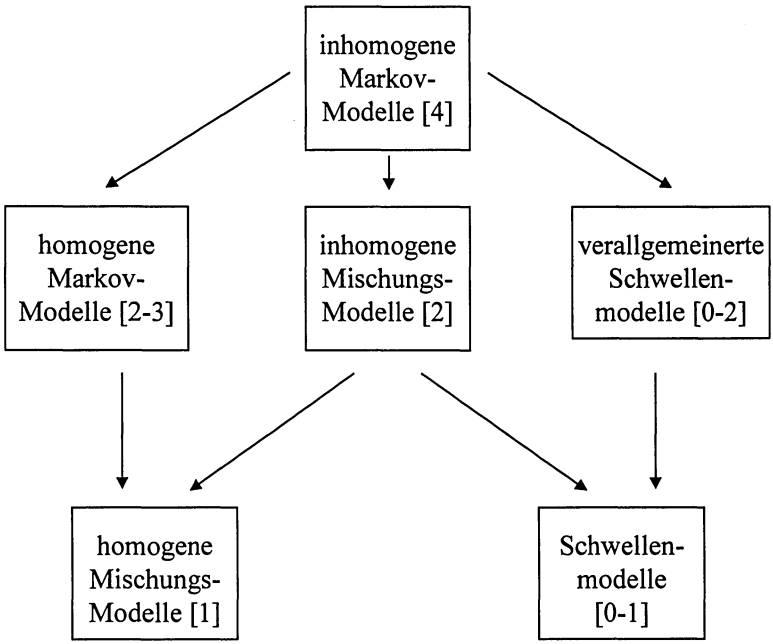

Abbildung 3: Systematik der Regimewechselmodelle. Zahlen in eckigen Klammern geben die Anzahl der freien Parameter des Zustandsprozesses s_t an.

2.4 Weitere Modelle mit veränderlichen Parametern

Außer Regimewechselmodellen sind in der ökonometrischen Literatur noch verschiedene weitere Modelltypen untersucht worden, die es ermöglichen, Parametervariation zu erkennen und zu modellieren.

Ein Überblick über die am weitesten verbreiteten Modellklassen findet sich z.B. in JUDGE u.a. [1985], S. 798, oder in PFAFF [1998], S. 37, wo die in dieser Arbeit untersuchten Regimewechselmodelle als *Modelle mit diskreter Koeffizientenvariation* bzw. als *switching regression models* bezeichnet werden. Regimewechselmodelle werden in diesen Klassifikationen zu den Modellen mit *systematischer Koeffizientenvariation* (*varying but nonstochastic parameters*) gezählt, was im Falle von Mischungs- und Markov-Modellen nicht ganz korrekt ist, da hier die Koeffizienten durchaus stochastisch variieren.[28] Schwellenmodelle, deren

[28] Unter den Modellen mit stochastischer Koeffizientenvariation wird üblicherweise noch unterschieden zwischen denen, deren Koeffizienten einem *stationären* Prozeß entstammt, und denen, deren Koeffizientenprozeß

Schwellenindikator verzögert endogen oder exogen und stochastisch ist, weisen ebenfalls stochastische Koeffizientenvariation auf.

Regimewechselmodelle decken also die ganze Palette der denkbaren Formen von Koeffizientenvariation ab. Was sie unter den Modellen mit veränderlichen Parametern auszeichnet, ist die Tatsache, daß die Parameter nur endlich vieler Werte fähig sind. Im Gegensatz dazu können die Parameter bei den sonstigen Modellen mit veränderlichen Parametern, den *kontinuierlichen* Modellen, theoretisch alle Werte annehmen.

Es ist instruktiv, die hier behandelten Regimewechselmodelle als Spezialfälle bekannter Modelle mit kontinuierlich variierenden Parametern aufzufassen. Dies wird im folgenden an zwei wichtigen Klassen demonstriert, den Modellen mit variablen Parametern im Sinne von RAJ und ULLAH [1981] und LÜDEKE, HUMMEL und RÜDEL [1989] sowie den Zustandsraummodellen.

2.4.1 Regimewechselmodelle als Modelle mit variablen Parametern

Aufbauend auf HILDRETH und HOUCK [1968] und RAJ und ULLAH [1981] wird in LÜDEKE, HUMMEL und RÜDEL [1989] folgendes Modell vorgeschlagen:[29]

$$y_t = \mathbf{W}_t\alpha + \mathbf{X}_t\beta_t + u_t$$

mit y_t, $\mathbf{W}_t$, $\mathbf{X}_t$, α und u_t wie oben[30] und

$$\beta_t = \mathbf{Z}_t\gamma + \varepsilon_t \ .$$

Die Matrix $\mathbf{Z}_t$ enthält die systematischen Bestimmungsgrößen von β_t,

$$\mathbf{Z}_t = \begin{pmatrix} z_{1,1,t} \cdots z_{1,n_1,t} & 0 \cdots 0 & \cdots & 0 \cdots 0 \\ 0 \cdots 0 & z_{2,1,t} \cdots z_{2,n_2,t} & \cdots & 0 \cdots 0 \\ \vdots \ \vdots & \vdots \ \vdots & \ddots & \vdots \ \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & z_{L,1,t} \cdots z_{L,n_L,t} \end{pmatrix},$$

und γ ist ein entsprechend dimensionierter Spaltenvektor. Dies ist ein Schwellenmodell mit Schwellenvariable z_t und Schwellenwert c, falls die Bestimmungsgrößen von β_t geeignete Dummy-Variablen darstellen und die stochastische Komponente von β_t nicht existiert, d.h. falls

$$n_1 = n_2 = \ldots = n_L = 2 \ ,$$

$$z_{i,1,t} = D_c(z_t) \text{ und } z_{i,2,t} = 1 - D_c(z_t) \text{ für } i = 1,\ldots,L \ ,$$ [31]

instationär ist. Homogene Mischungsmodelle und homogene, stationäre Markov-Modelle gehören zu den ersteren, die übrigen Mischungs- und Markov-Modelle zu den letzteren (vgl. den Anhang zu diesem Kapitel).

[29] Die Notation wurde leicht der hier verwendeten angepaßt.

$$\gamma = (\beta_{1,1}, \beta_{2,1}, \beta_{1,2}, \beta_{2,2}, \ldots, \beta_{1,L}, \beta_{2,L})'$$

und $$\varepsilon_t = \mathbf{0}\,, t = 1, \ldots, T\ .$$

Wenn andererseits die systematische Komponente der Koeffizienten entfällt (d.h. $\mathbf{Z}_t = \mathbf{0}$) und ε_t nicht normalverteilt ist (wie es in LÜDEKE, HUMMEL und RÜDEL [1989] angenommen wird), sondern nur der beiden Werte β_1 und β_2 fähig ist, d.h.

$$\varepsilon_t = \beta_{s_t}\ ,\ t = 1, \ldots, T\ ,$$

wobei s_t eine Bernoullifolge bzw. eine Markov-Kette wie in Kapitel 2.2.1 ist, so liegt ein Mischungs- bzw. Markov-Modell vor.

2.4.2 Homogene Markov-Modelle als Zustandsraummodelle

Sei $\mathbf{e}_1 = (1,0)'$, $\mathbf{e}_2 = (0,1)'$ und $\xi_t = \mathbf{e}_{s_t}$. Dann gilt

$$E(\xi_t) = (P(s_t = 1), P(s_t = 2))'\ ,$$

$$E(\xi_{t+1} | s_t = i) = (p^{1i}, p^{2i})'\ ,$$

und allgemeiner $$E(\xi_{t+1} \mid \xi_t, \xi_{t-1}, \ldots, \xi_1) = \mathbf{P}\xi_t$$

mit der Übergangsmatrix

$$\mathbf{P} = \left(p^{ij}\right) = \begin{pmatrix} p^{11} & \cdots & p^{1n} \\ \vdots & & \vdots \\ p^{n1} & \cdots & p^{nn} \end{pmatrix}.$$

Aus den Eigenschaften der bedingten Erwartung folgt deswegen

$$\xi_{t+1} = \mathbf{P}\xi_t + \mathbf{v}_{t+1}\ ,$$

wobei $\mathbf{v}_{t+1}$ unkorreliert mit $\xi_1, \ldots, \xi_t$ ist, und ebenso mit den entsprechenden exogenen Größen, falls die Markov-Kette als unabhängig von ihnen angenommen wird.[32]

In dieser Form stellt die dynamische Spezifikation der Markov-Kette die Zustandsgleichung eines linearen Zustandsraummodells dar.[33] Die Beobachtungsgleichung erhält man, wenn man die Regressionsbeziehung in der Form

$$y_t = \mathbf{W}_t\alpha + \mathbf{X}_t\mathbf{B}\xi_t + u_t$$

---

[30] Die Variable u_t ist in LÜDEKE, HUMMEL und RÜDEL [1989] homoskedastisch.

[31] D_c ist die in Kapitel 2.1 eingeführte Stufenfunktion mit Sprung bei c.

[32] Dies ist eine übliche und sinnvolle Annahme für s_t. Die stochastischen Spezifikationen des Modells werden in Kapitel 4.1 präzisiert.

[33] Vg. hierzu HAMILTON [1994a], S. 3062 ff., oder KROLZIG [1997], S. 29 ff.

mit $\mathbf{B} = (\beta_1, \beta_2)$ darstellt. Formal stellt ein Markov-Modell damit ein Zustandsraummodell dar, aber mit der Besonderheit, daß der Störterm $\mathbf{v}_t$ nicht normalverteilt ist, sondern nur vier Werte annehmen kann. Deswegen sind die für Zustandsraummodelle (mit normalverteilten Störtermen) entwickelten Methoden auch nicht ohne weiteres auf Markov-Modelle übertragbar. Diese Verwandtschaft zwischen den Modellen erklärt jedoch viele Ähnlichkeiten zwischen den jeweiligen Analyseverfahren, insbesondere die Analogie zwischen dem Kalman-Filter und dem Filter zur Bestimmung der sogenannten *geglätteten Wahrscheinlichkeiten*, die Rückschlüsse auf die unbeobachteten Zustände eines Markov-Modells gestatten.[34]

Anhang: Markov-Ketten

In diesem Abschnitt werden die Eigenschaften von Markov-Ketten behandelt, die für die späteren Ausführungen benötigt werden.[35]

Ein stochastischer Prozeß s_t mit Zustandsraum $\{1, 2, 3,...\}$ heißt *Markov-Kette der Ordnung* p, wenn er die folgende *Markov'sche Eigenschaft* besitzt:

Für alle t, m ∈ ⌉ mit m ≥ p und $P(s_t|s_{t-1}, s_{t-2},..., s_{t-p}) > 0$ *gilt*

$$P(s_t|s_{t-1}, s_{t-2},..., s_{t-m}) = P(s_t|s_{t-1}, s_{t-2},..., s_{t-p}) \ .$$

Die Verteilung von s_t hängt also nur von den letzten p Realisationen des stochastischen Prozesses ab. Wenn im folgenden von Markov-Ketten (schlechthin) die Rede ist, so sind damit immer Markov-Ketten *erster* Ordnung gemeint. Für eine Markov-Kette s_t gilt demnach

$$P(s_t|s_{t-1}, s_{t-2},..., s_{t-m}) = P(s_t|s_{t-1})$$

für jedes $m \geq 1$. Die Dynamik von Markov-Ketten mit endlichem Zustandsraum $\{1, 2, ..., n\}$ wird also von den n^2 *Übergangswahrscheinlichkeiten*

$$p_t^{ij} = P(s_t = i|s_{t-1} = j) \ , \ 1 \leq i, j \leq n$$

bestimmt,[36] wobei $\sum_{i=1}^{n} p_t^{ij} = 1$ für $j = 1,..., n$ gilt. Die Übergangswahrscheinlichkeiten bilden die *Übergangsmatrix*

$$\mathbf{P}_t = \left(p_t^{ij}\right) = \begin{pmatrix} p_t^{11} & \cdots & p_t^{1n} \\ \vdots & & \vdots \\ p_t^{n1} & \cdots & p_t^{nn} \end{pmatrix} .$$

[34] Vgl. hierzu Kapitel 4.3.2.1 (insbesondere S. 56) und Anhang A zu Kapitel 4.

[35] Vgl. hierzu z.B. KARLIN und TAYLOR [1975] oder HAMILTON [1994b], Kap. 22.2.

[36] Einige Autoren verwenden die Konvention $p^{ij} = P(s_t = j \mid s_{t-1} = i)$.

Aus der Markov'schen Eigenschaft erhält man für $1 \le i, k \le n$

$$P(s_t = i | s_{t-2} = k) = \sum_{j=1}^{n} P(s_t = i | s_{t-1} = j) P(s_{t-1} = j | s_{t-2} = k) = \left(\mathbf{P}_t \mathbf{P}_{t-1}\right)_{ik} .$$

Daraus resultiert durch wiederholte Anwendung die sogenannte *Chapman-Kolmogorov-Gleichung*

$$P(s_t = i | s_{t-r} = k) = \left(\mathbf{P}_t \mathbf{P}_{t-1} \cdots \mathbf{P}_{t-r+1}\right)_{ik} .$$

Ist $n = 2$, so gilt

$$\mathbf{P}_t = \begin{pmatrix} p_t^{11} & p_t^{12} \\ p_t^{21} & p_t^{22} \end{pmatrix} = \begin{pmatrix} p_t^{11} & 1 - p_t^{22} \\ 1 - p_t^{11} & p_t^{22} \end{pmatrix} ,$$

wenn man p_t^{11} und p_t^{22} als freie Parameter verwendet.

Die Markov-Kette heißt *homogen* oder *Kette mit stationären Übergangswahrscheinlichkeiten*, falls $\mathbf{P}_t$ unabhängig von t ist. Für $n = 2$ gilt dann

$$\mathbf{P}_t \equiv \mathbf{P} = \begin{pmatrix} p^{11} & 1 - p^{22} \\ 1 - p^{11} & p^{22} \end{pmatrix} .$$

Eine Markov-Kette heißt *stationär*, falls die *unbedingten* Verteilungen der s_t zeitlich konstant sind, d.h.

$$P(s_t = i) \equiv \pi_i \quad \text{für alle } t.$$

Für eine homogene, stationäre Markov-Kette gilt

$$\pi_i = P(s_t = i) = \sum_{j=1,2} P(s_t = i | s_{t-1} = j) P(s_{t-1} = j) = \sum_{j=1,2} p^{ij} \pi_j ,$$

d.h. $$\pi = \mathbf{P}\pi$$

mit $\pi = (\pi_1, \ldots, \pi_n)'$. Der Vektor der unbedingten Verteilung von s_t ist also ein Eigenvektor der Übergangsmatrix zum Eigenwert 1. Für $n = 2$ lauten die Komponenten dieses Eigenvektors (der durch die Normierung $\pi_1 + \pi_2 = 1$ festgelegt ist)

$$\pi_1 = \frac{1 - p^{22}}{(1 - p^{11}) - (1 - p^{22})}, \pi_2 = 1 - \pi_1 = \frac{1 - p^{11}}{(1 - p^{11}) - (1 - p^{22})} .$$

Diese Größen werden auch als *unbedingte Wahrscheinlichkeiten* für die beiden Zustände bezeichnet.

Die Verteilung einer inhomogenen Markov-Kette s_t mit $t = 1,2,\ldots$ ist für $n = 2$ vollständig spezifiziert durch die Übergangswahrscheinlichkeiten p_t^{11} und p_t^{22} sowie die *Startwahrscheinlichkeit* $\rho^1 = P(s_1 = 1)$. Ist die Markov-Kette homogen, so wird sie durch die drei

Parameter p^{11}, p^{22} und ρ^1 charakterisiert: Es gilt dann aufgrund der Chapman-Kolmogorov-Gleichung für $t = 2,3,...$

$$\begin{pmatrix} P(s_t = 1) \\ P(s_t = 2) \end{pmatrix} = \mathbf{P}_t \mathbf{P}_{t-1} \cdots \mathbf{P}_2 \begin{pmatrix} \rho^1 \\ 1-\rho^1 \end{pmatrix},$$

bzw. im Falle einer homogenen Kette

$$\begin{pmatrix} P(s_t = 1) \\ P(s_t = 2) \end{pmatrix} = \mathbf{P}^{t-1} \begin{pmatrix} \rho^1 \\ 1-\rho^1 \end{pmatrix}.$$

Ist s_t stationär, so gilt zusätzlich die Restriktion

$$\rho^1 = \pi_1 = \frac{1-p^{22}}{(1-p^{11})-(1-p^{22})}.$$

Die Verteilung der Markov-Kette wird in diesem Fall von zwei freien Parametern beschrieben.

Die unbedingten Wahrscheinlichkeiten stellen die Verteilung (fast) jeder Markov-Kette für große t dar, auch wenn sie nicht strikt stationär ist, etwa weil die Startwahrscheinlichkeiten nicht die unbedingten Wahrscheinlichkeiten sind. Genauer: Sind alle Übergangswahrscheinlichkeiten p^{ij} positiv, so konvergieren die t-Schritt-Übergangswahrscheinlichkeiten $(\mathbf{P}^t)_{ij}$ für wachsende t exponentiell gegen die unbedingten Wahrscheinlichkeiten π_i. Da diese nicht von j abhängen, stellen sie also die Wahrscheinlichkeiten für die Zustände für große t dar.

Ein wichtiger Spezialfall liegt vor, wenn $p_t^{11} = p_t^{12}$ für $t = 2, 3,...$ gilt, also $p_t^{11} = 1 - p_t^{22}$. Dann gilt auch $p_t^{22} = p_t^{21}$, und das bedeutet

$$P(s_t | s_{t-1}) = P(s_t).$$

In diesem Fall sind die einzelnen Realisationen der Markov-Kette unabhängig bernoulliverteilt mit Parameter $p_t = p_t^{11}$. Ist s_t homogen, so gilt $p_t \equiv p^{11}$, und der Prozeß s_t bildet eine Bernoullifolge mit dem Parameter $p = p^{11}$.

Gilt $s_t = 1$, so ist die Wahrscheinlichkeit dafür, daß die Kette noch genau $n-1$ (also insgesamt n) Zeitpunkte in diesem Zustand verbleibt, gleich $(p^{11})^{n-1}(1-p^{11})$. Die *Verweildauer* in diesem Zustand ist also geometrisch verteilt mit Erwartungswert $1/(1-p^{11})$.

3 Wechselkurse im Europäischen Währungssystem

Die Wechselkurse im Europäischen Währungssystem (EWS) wurden maßgeblich durch die institutionellen Rahmenbedingungen des EWS bestimmt. Für die Modellierung dieser Wechselkurse ist deshalb eine Analyse der Funktionsweise des EWS erforderlich, die sich in Kapitel 3.1 zusammen mit einem kurzen Abriß der historischen Entwicklung des EWS und einem Ausblick auf das Nachfolgesystem des EWS nach Inkrafttreten der Europäischen Währungsunion (EWU) findet. In Kapitel 3.2 werden die theoretischen Grundlagen von Wechselkursmodellen dargestellt. In Kapitel 3.3 schließlich wird der in dieser Arbeit verwendete Ansatz der Modellierung von Wechselkursdaten mit Hilfe von Regimewechselmodellen erläutert.

3.1 Das Europäische Währungssystem

3.1.1 Geschichtlicher Hintergrund und Entstehung des EWS

Nach dem endgültigen Zusammenbruch des Bretton-Woods-Systems 1973 unternahmen die europäischen Zentralbanken und Regierungen bald die ersten Versuche, wenigstens innerhalb Europas ein gewisses Maß an Wechselkursstabilität zu garantieren.[37] Diesen frühen, unter der Bezeichnung „Europäische Währungsschlange" bekannten Initiativen war durchweg wenig Erfolg beschieden. Die festgelegten Paritäten mußten regelmäßig nach einigen Monaten wieder aufgegeben werden, so daß sie nie wirklich das Vertrauen der Märkte genossen. Dennoch wurde das Ziel fester Wechselkurse vor allem innerhalb der Europäischen Union (EU, damals Europäische Gemeinschaft, EG) von den verantwortlichen Regierungs- und Zentralbankpolitikern als grundlegend für die ökonomische Integration und insbesondere für das Funktionieren des gemeinsamen Agrarmarktes[38] angesehen, so daß weiterhin Möglichkeiten für die Wiederherstellung eines höheren Maßes an Wechselkursstabilität erörtert wurden.

Als Hauptgrund für das Scheitern der Währungsschlange kristallisierte sich das Fehlen von verbindlichen, institutionalisierten Interventionsregeln der Zentralbanken heraus. Auf Tagungen in Kopenhagen und Bremen beschloß der Europäische Rat im Jahr 1978 die Grundzüge einer neuen Währungsordnung mit solchen Interventionsverpflichtungen, so daß am 13. März

[37] Vgl. hierzu und zu den weiteren Ausführungen über Funktionsweise und ökonomischer Bedeutung des EWS z.B. GIAVAZZI, MICOSSI und MILLER [1988] sowie GIAVAZZI und GIOVANNINI [1989].

[38] Vgl. hierzu GIAVAZZI und GIOVANNINI [1989], S. 12 ff.

1979 das EWS unter Teilnahme von Belgien, Dänemark, der Bundesrepublik Deutschland, Frankreich, Irland, Italien, Luxemburg und den Niederlanden in Kraft treten konnte.[39]

3.1.2 Die Funktionsweise des EWS

3.1.2.1 Der ECU und das Paritätengitter

Den ersten Schritt in der Konstruktion des EWS stellte die Einführung der Europäischen Währungseinheit (*European Currency Unit*, ECU) dar.[40] Der ECU war eine Korbwährung, zusammengesetzt aus einer festen Anzahl von Währungseinheiten aller EU-Währungen. Diese Festlegung führte zum einen dazu, daß die EWS-Teilnehmerstaaten nicht identisch mit den im ECU repräsentierten Währungen waren, und zum anderen dazu, daß sich die Gewichte der einzelnen Währungen im ECU, die ursprünglich auf der Höhe der Bruttosozialprodukte der einzelnen Länder basierten, im Laufe der Zeit infolge von Wechselkursveränderungen mitunter erheblich veränderten. Zudem wurde die Zusammensetzung des ECU in den Jahren 1984 und 1989 verändert.

Die wichtigsten Funktionen des ECU waren:

- Er diente als Reservemedium und als Recheneinheit zum Saldenausgleich zwischen den EWS-Zentralbanken, ähnlich wie das Sonderziehungsrecht beim Internationalen Währungsfonds.
- Er stellte die Basiseinheit der ECU-Leitkurse dar.

Diese von den Zentralbanken festgelegten *ECU-Leitkurse*, ausgedrückt in den Preisen eines ECU in den nationalen Währungen, waren das Kernstück des EWS. Die Quotienten der ECU-Leitkurse bildeten die *bilateralen Leitkurse* der jeweiligen Staaten und stellten zusammen das *Paritätengitter* dar.

Ursprünglich sollte der ECU eine ähnlich zentrale Rolle im EWS einnehmen wie der Dollar im Bretton-Woods-System. Dieser Standpunkt wurde in der Frühzeit des EWS besonders von Frankreich vertreten. Tatsächlich setzte sich aber, wie im folgenden Kapitel erläutert wird, die Bundesbank mit ihrer Forderung durch, den Interventionsmechanismus an das Paritätengitter zu binden, wodurch der ECU vorwiegend zu einer Rechengröße der Zentralbanken entwertet wurde. Dabei mag auch eine Rolle gespielt haben, daß der ECU aufgrund seiner oben erläu-

[39] Aufbau, Arbeitsweise und Entwicklung des EWS werden in den Monatsberichten der Bundesbank vom März 1979, November 1989, Oktober 1992 und August 1993 detailliert erläutert.

[40] „ECU" ist nicht nur die oben erläuterte Abkürzung, sondern auch der Name einer alten französischen Goldmünze, wodurch sich die übliche (französische) Aussprache erklärt.

terten Konstruktion teilweise komplizierte Umrechnungsverfahren nötig machte, was ihn für praktische Zwecke ungeeignet erscheinen ließ.[41]

3.1.2.2 Der Interventionsmechanismus

Wie bereits erwähnt hatte Frankreich gefordert, den Interventionsmechanismus an Abweichungen der einzelnen Währungen vom ECU auszurichten. Dies scheiterte aber vorwiegend am Widerstand der Bundesbank, die befürchtete, aufgrund der erwarteten Stärke der DM im EWS zu häufig zu einseitigen Interventionen gezwungen zu werden, was zu einer unerwünschten Ausweitung der (DM-)Geldmenge und damit zu einer Gefährdung der Stabilität der DM hätte führen können. Statt dessen setzte sie sich mit ihrem Vorschlag durch, die Interventionsverpflichtungen an das bilaterale Paritätengitter zu koppeln, so daß wenigstens immer eine zweite Zentralbank zu entgegengesetzten Interventionen verpflichtet war.

Der Interventionsmechanismus sah vor, daß sich die beteiligten Zentralbanken verpflichteten, durch Interventionen auf den Devisenmärkten die bilateralen Wechselkurse in einem festen Band um die bilateralen Leitkurse zu halten. Die Ober- und Untergrenze dieses Bandes lagen anfangs für die meisten Währungen 2,25 % ober- bzw. unterhalb der bilateralen Leitkurse, so daß die Bandbreite 4,5 % betrug. Ausnahmen stellten die italienische Lira und später die spanische Peseta sowie der portugiesische Escudo mit Bandbreiten von 12 % dar. Am 2. August 1993 wurden diese Bandbreiten aufgrund des anhaltenden spekulativen Drucks auf den Devisenmärkten nach der „Krise des EWS“ (vgl. Kapitel 3.1.2.4) auf 30 % erhöht, mit Ausnahme der DM und des Gulden, für die aufgrund einer bilateralen Übereinkunft eine Bandbreite von 4,5 % beibehalten wurde.

Sobald der Wechselkurs zweier beteiligter Länder eine dieser Grenzen erreichte, waren beide Zentralbanken verpflichtet, die jeweils schwächere Währung zu kaufen und die stärkere zu verkaufen, und zwar in unbeschränktem Umfang. Um der Zentralbank der schwächeren Währung diese Interventionen zu ermöglichen, war die Zentralbank der stärkeren Währung verpflichtet, ihr unbegrenzten Kredit zu gewähren. Dieser Kreditmechanismus und die zugehörige Verzinsung wurde in den verschiedenen Formen der finanziellen Beistandsmechanismen genau festgelegt.[42]

Außer diesen sogenannten *marginalen Interventionen* spielten die *intramarginalen Interventionen* eine wesentliche Rolle, also Eingriffe der beteiligten Zentralbanken in die Devisenmärkte, bevor die betroffenen Währungen eine der Grenzen der Wechselkursbänder er-

[41] Vgl. hierzu z.B. GIAVAZZI und GIOVANNINI [1989], S. 35 f.

[42] Vgl. hierzu den Monatsbericht der Bundesbank vom März 1979, S. 15 ff.

reichten. Für solche Interventionen, die der Zustimmung der jeweils anderen Zentralbank bedurften, war jedoch kein Kreditmechanismus vorgesehen. Die beteiligten Zentralbanken engagierten sich in unterschiedlichem Ausmaß in intramarginalen Interventionen: Während z.B. die Banque de France eine sehr aktive Rolle in diesem Zusammenhang einnahm, beteiligte sich die Bundesbank prinzipiell nicht an solchen Eingriffen.[43]

3.1.2.3 Das Verfahren zur Neufestlegung der Paritäten

Obwohl mit den unbegrenzten Interventionsverpflichtungen der beteiligten Zentralbanken das größtmögliche Vertrauen der Märkte in die Leitkurse sichergestellt werden sollte, waren *Realignments*, also Veränderungen der Leitkurse, von Anfang an vorgesehen und wurden auch von Zeit zu Zeit nötig. Der Hauptgrund dafür waren die zum Teil erheblich von einander abweichenden Inflationsraten in den Teilnehmerländern und der mit diesem Abweichen von dem durch die Kaufkraftparitätentheorie vorgegebenen Gleichgewichtspfad[44] einhergehende Ab- bzw. Aufwertungsdruck auf die schwächere bzw. die stärkere Währung auf den Devisenmärkten.

Wichtigstes Merkmal des zugehörigen Verfahrens war die Notwendigkeit der Zustimmung *aller* beteiligten Länder zu neuen Paritäten. Dies stellte einen Hauptunterschied zum System von Bretton-Woods und zur Währungsschlange dar, in denen Realignments letztlich einseitige Entscheidungen des jeweiligen Landes darstellten. Nach erfolgter Zustimmung traten die neuen Leitkurse ohne Vorankündigung sofort in Kraft.

3.1.2.4 Die Entwicklung des EWS von 1979 bis zur EWU

Obwohl in der Konstruktion des EWS bewußt die Bezugnahme auf eine (echte) Leitwährung vermieden worden war, entwickelte sich die DM im Verlaufe des Bestehens des EWS zur Ankerwährung und nahm zeitweise eine ähnliche Stellung ein wie der Dollar im Bretton-Woods-System. Dies zeigte sich vor allem daran, daß Interventionen vorwiegend dann nötig wurden, wenn verschiedene Währungen ihr Wechselkursband zur DM zu verlassen drohten, woraus in der Regel auch die Notwendigkeit von Realignments erwuchs. Faktisch entschied die Bundesbank aus stabilitätsorientierten Überlegungen über ihre Geldpolitik; die übrigen Mitgliedsstaaten hatten die Wahl, entweder diesem Kurs in ihrer Geldpolitik zu folgen, wie es vorwiegend in der jüngeren Zeit der Fall gewesen zu sein scheint, oder ihre Währungen gegenüber der DM abzuwerten, wie es in der ersten Zeit des EWS regelmäßig vorkam.

---

[43] Vgl. GIAVAZZI und GIOVANNINI [1989], S. 65.

[44] Vgl. hierzu Kapitel 3.2.1.1.

Realignments hatten unterschiedliche Ausmaße, von der Veränderung eines einzelnen bilateralen Leitkurses (wie z.B. der Abwertung der Dänischen Krone im November 1979) bis zur kompletten Umgestaltung des Paritätengitters (wie z.B. dem Realignment vom Oktober 1981). Insgesamt läßt sich die Dauer des EWS in vier Phasen einteilen:

- die ersten Jahre bis zur Konsolidierung des EWS etwa März 1983, mit relativ vielen Realignments (insgesamt 7)
- die vergleichsweise ruhige Phase bis Mitte 1992, in der nur 3 Realignments nötig waren und Spanien im Juni 1989 sowie Portugal im April 1992 dem EWS beitraten
- die Krise des EWS bis zur Erweiterung der Bandbreiten auf 30 % am 2. August 1993, auf deren Höhepunkt Großbritannien und Italien im September 1992 aus dem EWS ausschieden, nachdem sie praktisch ihre gesamten Devisenreserven aufgebraucht hatten, um offensichtlich überhöhte Wechselkurse aus politischen Gründen zu verteidigen
- die erneute Konsolidierung bis zum Beginn der Europäischen Währungsunion am 1.1.1999, mit relativ stabilen Wechselkursen trotz hoher Bandbreiten, dem Beitritt Österreichs (Januar 1995) und Finnlands (Oktober 1996) und dem Wiedereintritt Italiens (November 1996).

3.1.2.5 Der Wechselkursmechanismus seit dem 1. Januar 1999

Mit Beginn der Europäischen Währungsunion endete das EWS, während gleichzeitig der Euro als echte Gemeinschaftswährung den ECU im Verhältnis 1:1 ablöste. Da der Maastrichter Vertrag die zweijährige abwertungsfreie Teilnahme am EWS zur Voraussetzung für den Beitritt eines Landes zum Euro macht, jedoch (noch) nicht alle EU-Staaten diesen Schritt unternommen haben (und mittelfristig mit dem Beitritt mehrerer zentraleuropäischer Länder zur EU zu rechnen ist), wurde im September 1998 von der Europäischen Zentralbank (EZB) die Ausgestaltung des früher als EWS II und heute von der Bundesbank als WKM II (Wechselkursmechanismus II) bezeichneten Nachfolgesystems des EWS beschlossen und ein entsprechendes Abkommen von der EZB und den 4 bisher nicht dem Euro beigetretenen EU-Ländern – Dänemark, Griechenland, Großbritannien und Schweden – unterzeichnet.[45] Dieses Abkommen stellt den nicht zum Euro übergegangenen EU-Mitgliedern frei, wann sie dem WKM II beitreten. Dänemark und Griechenland haben dies zum 1.1.1999 getan.

Die Konstruktion des WKM II ähnelt stark der des EWS. Es bestehen jedoch einige bedeutsame Unterschiede, die im folgenden kurz zusammengefaßt werden.

[45] Vgl. hierzu den Monatsbericht der Bundesbank vom Oktober 1998.

- Das Paritätengitter wird von den Leitkursen zwischen den betroffenen Währungen und dem Euro, der also die Ankerwährung im WKM II wird und damit die informelle Rolle der DM im EWS offiziell übernimmt, abgelöst.
- Auch die Bandbreiten werden nur noch im Verhältnis zum Euro bestimmt. Sie betragen im Normalfall 30 % (wie z.Zt. für die Griechische Drachme), können aber auf Antrag des betroffenen Mitgliedslandes auch reduziert werden (für die Dänische Krone z.Zt. auf 4,5 %).
- Die unbegrenzte Interventionsverpflichtung der EZB und der betroffenen nationalen Zentralbank kann ausgesetzt werden, wenn durch diese Eingriffe die Preisstabilität im Euro-Raum gefährdet werden könnte.
- Die EZB beteiligt sich grundsätzlich nicht an intramarginalen Interventionen.
- Die betroffenen Zentralbanken erhalten für marginale Interventionen prinzipiell unbegrenzten Kredit bei der EZB, für intramarginale Interventionen jedoch nur bis zum Erreichen einer festgelegten Kreditobergrenze.

Zusammenfassend läßt sich zum einen sagen, daß mit der Bindung an den Euro die Fiktion eines echten multilateralen Systems aufgegeben wird, das faktisch auch im EWS nicht bestanden hat, und so die reale Existenz einer Ankerwährung offiziell akzeptiert wird. Damit nähert sich der WKM II formal wieder dem System von Bretton-Woods an.

Zum anderen ist die Verantwortung für das Verbleiben der betroffenen Währungen im WKM II nicht gleichmäßig zwischen der EZB und den betroffenen Zentralbanken aufgeteilt. Die EZB sichert ihren Beistand zur Überwindung spekulativer Krisen, die nicht durch Fundamentaldaten begründet sind, zu, überläßt die übrigen Maßnahmen (etwa intramarginale Interventionen) aber vorwiegend den betroffenen Mitgliedsstaaten. Ähnlich hatte sich ja bereits die Bundesbank im EWS verhalten. Insbesondere behält sich die EZB ausdrücklich die Option vor, aus stabilitätspolitischen Gründen die Stützung einer unter Druck geratenen Währung einzustellen.

3.2 Theoretische Grundlagen für Wechselkursmodelle

Im ersten Abschnitt dieses Kapitels (3.2.1) werden die Elemente monetärer Wechselkursmodelle erläutert, die für die empirischen Untersuchungen benötigt werden. Ihre Integration in Gesamtmodelle mit flexiblen Wechselkursen wird in Abschnitt 3.2.2 nur kurz behandelt, da auf diese Modelle im weiteren nur an wenigen Stellen Bezug genommen wird.

Im folgenden Abschnitt 3.2.3 werden theoretische Ansätze zur Erklärung von Wechselkursen in Bandbreitensystemen behandelt. Nach einer kurzen Würdigung der Modelle vom Krugmann-Typ (die auf den Ansatz von KRUGMANN [1991] zurückgehen) wird genauer auf Modellansätze eingegangen, die die Besonderheiten des EWS berücksichtigen.

3.2.1 Determinanten des Wechselkurses

3.2.1.1 Kaufkraftparitätentheorie

Den wohl wichtigsten und ältesten Ansatz zur Erklärung des Verlaufs von Wechselkursen zwischen zwei Ländern (dem „Inland" und dem „Ausland") stellt die *Kaufkraftparitätentheorie* (*Purchasing Power Parity Theory*, PPP) dar.[46]

Kerngedanke der PPP ist, daß grenzüberschreitende Arbitrageprozesse dazu führen, daß der inländische und der mit dem Wechselkurs auf die Inlandswährung umgerechnete ausländische Preis international handelbarer Güter sich gleichen müssen. Liegen Preisdifferenzen vor, die (unter Berücksichtigung von Transport-, Geldwechselkosten u.ä.) Arbitrage lohnend machen, so werden diese Güter verstärkt in dem Land mit dem niedrigeren Preisniveau gekauft und in das jeweils andere Land im- bzw. exportiert. Durch diese Nachfrageverlagerung steigen die Güterpreise in dem Land mit dem niedrigeren Preisniveau und sinken in dem anderen. Außerdem hat der Güterstrom einen Geldstrom in die umgekehrte Richtung zu Folge. Die damit einhergehenden Angebots- bzw. Nachfrageveränderungen auf dem Devisenmarkt führen zu einem Anstieg des Wechselkurses der Währung des Landes mit dem niedrigeren Preisniveau. Beide Effekte führen zu einem Angleichen der Güterpreise im internationalen Vergleich, d.h. zu einer Veränderung der nationalen Güterpreise und des Wechselkurses in Richtung auf das von der PPP vorhergesagte Niveau.

Wird der Wechselkurs zum Zeitpunkt t, also der Preis der inländischen Währung in der ausländischen, mit w_t bezeichnet,[47] sowie das Güterpreisniveau im Inland zu diesem Zeitpunkt mit P_t und das im Ausland mit P_t^*, so bedeutet die Gültigkeit der PPP also, daß gilt

$$w_t = \frac{P_t^*}{P_t} .$$

[46] Zu den folgenden Ausführungen vgl. z.B. JARCHOW [1994], S. 214 ff., oder ROSE und SAUERNHEIMER [1995], S. 177 ff.

[47] In der deutschsprachigen Literatur ist – im Gegensatz zu der hier verwendeten *Mengennotierung* – die *Preisnotierung* weiter verbreitet, in der als Wechselkurs der Preis der ausländischen Währung in der inländischen bezeichnet wird. Diese Arbeit verwendet jedoch im Einklang mit der internationalen Literatur die Mengennotierung, die auch am 1.1.1999 an den europäischen Devisenbörsen für die Notierungen des Euro gegenüber anderen Währungen eingeführt wurde.

Diese sogenannte *absolute Form* der PPP impliziert für den realen Wechselkurs $w_t^r = w_t \cdot P_t / P_t^*$

$$w_t^r \equiv 1 \ .$$

Der reale Wechselkurs ist folglich zeitlich konstant und gleich eins. Weniger restriktiv als die absolute ist die *relative Form* der PPP, die man aus der absoluten durch Logarithmierung und Bildung der ersten zeitlichen Differenzen erhält. Sie besagt, daß gilt

$$\log(w_t / w_{t-1}) = \log(P_t^* / P_{t-1}^*) - \log(P_t / P_{t-1}) \ ,$$

oder (da für jede Variable x_t gilt: $\log(x_t / x_{t-1}) \approx \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{\Delta x_t}{x_{t-1}}$)

$$\frac{\Delta w_t}{w_{t-1}} \approx \frac{\Delta P_t^*}{P_{t-1}^*} - \frac{\Delta P_t}{P_{t-1}} \ .$$

Demnach entspricht die Veränderungsrate des Wechselkurses (näherungsweise) der Differenz zwischen der in- und der ausländische Güterteuerungsrate. Die relative Form der PPP ist äquivalent dazu, daß für den realen Wechselkurs gilt

$$w_t^r = w_{t-1}^r \ ,$$

d.h.

$$w_t^r = k$$

mit einer Konstanten k. Der reale Wechselkurs ist also zeitlich konstant; die absolute PPP erhält man als Spezialfall für $k = 1$.

Häufig wird die Gültigkeit der PPP nicht nur für einzelne Güter gefordert, sondern für einen ganzen Warenkorb international handelbarer Güter. Für diesen Warenkorb kann die PPP i.a. natürlich nicht mehr exakt, sondern nur noch näherungsweise gelten. Gilt die relative Form der PPP für einen solchen Warenkorb, so kann man die in- und ausländischen Preisveränderungsraten natürlich auch mit Hilfe passender Preis*indizes* berechnen. In empirischen Anwendungen wird hier gewöhnlich der Verbraucherpreisindex (*Consumer Price Index*, CPI) des jeweiligen Landes verwendet.[48]

3.2.1.2 Zinsparitätentheorie

Das finanzmarkttheoretische Äquivalent zur PPP ist die Zinsparitätentheorie. Anstelle von (realen) Gütern werden bei der Zinsparitätentheorie zinstragende in- und ausländische Wertpapiere betrachtet. Als zusätzlicher Aspekt kommt hinzu, daß die Laufzeit der Anlage und

[48] Alternativ hierzu bietet sich der Preisindex für gewerbliche Erzeugnisse (*Wholesale Price Index*, WPI) an. Einige Autoren argumentieren, daß sich darin die international handelbaren Güter genauer widerspiegeln, eine Meinung, die aber nicht von allen Autoren geteilt wird. Vgl. hierzu z.B. DREGER [1996], S. 50, oder DIEBOLD [1988], S. 117.

damit eventuelle Wechselkursveränderungen bis zum Auszahlungstermin berücksichtigt werden müssen. Dies kann je nach Anlagestrategie in unterschiedlicher Weise erfolgen.

Bei der *gedeckten Zinsparität* (*Covered Interest Parity*, CIP) wird das Wechselkursrisiko durch eine Absicherung am Terminmarkt ausgeschaltet. Bezeichnet wieder w_t den Preis der ausländischen Währung in der inländischen, $f_{t,t+1}$ den Terminkurs zum Zeitpunkt t für den Zeitpunkt $t+1$, i_t den zeitstetig angepaßten nominellen Zinssatz im Inland mit Laufzeit von t bis $t+1$ und i_t^* den entsprechenden Zinssatz im Ausland, so besagt die CIP, daß gelten muß

$$e^{i_t} = \frac{w_t}{f_{t,t+1}} e^{i_t^*} .$$

Die linke Seite dieser Gleichung stellt den Auszahlungsbetrag zum Zeitpunkt $t+1$ für die Anlage einer Einheit der inländischen Währung zum Zeitpunkt t dar (oder die Kosten einer Kreditaufnahme in dieser Höhe, wenn man davon ausgeht, daß Kredit- und Anlagezinssatz übereinstimmen). Die rechte Seite repräsentiert den Auszahlungsbetrag zu diesem Zeitpunkt, wenn der Betrag statt dessen in die ausländische Währung getauscht, dort angelegt und gleichzeitig der Rücktausch in die inländische Währung zum Zeitpunkt $t+1$ durch ein Termingeschäft bereits zum Zeitpunkt t fixiert wird. Wird diese Anlage durch einen Kredit in der inländischen Währung zum Zinssatz i_t finanziert, so ist dieses Geschäft risikolos, und jede Abweichung von der CIP eröffnet ebensolche risikolosen Arbitragemöglichkeiten. Dies führt dazu, daß die CIP unter Berücksichtigung von Transaktionskosten zumindest für Anlagen an Euro-Kapitalmärkten (also außerhalb des Geltungsbereichs der betroffenen Währungen) in der Regel erfüllt ist.[49]

Bei der *ungedeckten Zinsparität* (*Uncovered Interest Parity*, UIP) wird auf die Absicherung durch das Termingeschäft verzichtet. Die Zinsdifferenz zwischen In- und Ausland ist dann ein Ausdruck für Erwartungen über die Entwicklung des Wechselkurses bis zum Auszahlungszeitpunkt. Bezeichnet $w_{t,t+1}^e$ den zum Zeitpunkt t für den Zeitpunkt $t+1$ erwarteten Wechselkurs, so bedeutet dies

$$e^{i_t} = \frac{w_t}{w_{t,t+1}^e} e^{i_t^*} ,$$

oder in relativer Form

$$i_t^* - i_t = \log(w_{t,t+1}^e / w_t) \approx \frac{w_{t,t+1}^e - w_t}{w_t} .$$

[49] Infolge von Kapitalverkehrskontrollen kann es sein, daß inländische Zinsen nicht marktgerecht sind und deshalb für sie Abweichungen von der CIP auftreten. Vgl. JARCHOW [1994], S. 244.

Die Zinsdifferenz zwischen Aus- und Inland ist also (näherungsweise) gleich der erwarteten Veränderungsrate des Wechselkurses.

3.2.2 Monetäre Wechselkursmodelle für flexible Wechselkurse

In den 70er Jahren wurden Kaufkraft- und Zinsparitätentheorie in monetaristische Gesamtmodelle integriert, in denen das Preisniveau von den jeweiligen Geldmärkten bestimmt wird.[50] Als zentrale Verhaltensgleichungen kommen in diesen Modellen die Geldnachfragefunktionen der beiden Länder hinzu. Weiter wird angenommen, daß die Erwartungsbildung rational ist, d.h.

$$w^e_{t,t+1} = E(w_{t+1} \mid I_t) \ ,$$

wobei I_t die Informationsmenge zum Zeitpunkt t bezeichnet. Unter diesen Annahmen werden Veränderungen des Geldangebotes quantitätstheoretisch direkt in entsprechende Veränderungen des Preisniveaus umgesetzt und führen über die PPP zu proportionalen Wechselkursveränderungen.

Die Annahme, daß die PPP auch kurzfristig jederzeit erfüllt ist, wurde jedoch von empirischen Studien in der Regel nicht bestätigt. Sie wurde deshalb in Modellen mit verzögerter Preisanpassung aufgegeben, bei denen die Güterpreise nicht mehr völlig flexibel sind.[51] In diesen Modellen bestimmt die PPP nur noch den langfristigen Wechselkurs. Die Zinsparitätentheorie hingegen ist auch kurzfristig immer erfüllt und stellt (über die Annahme rationaler Erwartungen) die Verbindung zum aktuellen Wechselkurs her. Die wichtigste Modellaussage ist, daß Veränderungen in der Geldbasis langfristig zu proportionalen Veränderungen in den Preisniveaus und dem Wechselkurs führen. Kurzfristig kann die Veränderung des laufenden Wechselkurses aufgrund der verzögerten Preisanpassung jedoch stärker sein (sogenannte *überschießende Wechselkursreaktionen*): Eine Ausweitung der inländischen Geldbasis führt sofort zu einem niedrigeren inländischen Zins und damit zu Kapitalexporten. Die damit einhergehende höhere Nachfrage nach Devisen kann den laufenden Wechselkurs über das langfristige Gleichgewichtsniveau hinaus erhöhen.

[50] Vgl. hierzu z.B. FRENKEL [1976].

[51] Vgl. hierzu DORNBUSCH [1976].

3.2.3 Modelle für Wechselkurse in Bandbreitensystemen

3.2.3.1 Modelle vom Krugmann-Typ

Von KRUGMANN [1991] stammt die erste Einbettung eines monetären Wechselkursmodells in einen „absolut glaubwürdigen" Zielzonen-Wechselkursmechanismus, d.h. ein Bandbreitensystem, bei dem die Wahrscheinlichkeit für ein Realignment gleich Null ist.[52] Sein bahnbrechendes Modell inspirierte eine Zahl weiterer Arbeiten, die Erweiterungen in verschiedener Hinsicht darstellen.[53] Allen diesen Modellen war aber insbesondere an EWS-Wechselkursdaten, also dem wesentlich wichtigeren Fall nicht absolut glaubwürdiger Zielzonen, nur sehr begrenzter empirischer Erfolg beschieden.[54] Aus diesem Grund und weil die Modelle in stetiger Zeit formuliert sind, also den Apparat der stochastischen Analysis benötigen, wird auf ihre Darstellung hier verzichtet.

3.2.3.2 Modelle für den Verlauf von Wechselkursen im EWS

Die in Kapitel 3.1 erläuterten Konstruktionselemente des EWS haben Konsequenzen für den zu erwartenden Verlauf der entsprechenden Wechselkurse, die von verschiedenen, nicht auf den Ansatz von Krugmann zurückgehenden Modellen für EWS-Wechselkurse aufgegriffen werden.

Zum einen ist damit zu rechnen, daß der Einfluß von Fundamentalfaktoren (etwa der Preis- oder der Geldmengenentwicklung) auf den Wechselkurs durch Interventionen der beteiligten Zentralbanken gedämpft wird. Wenn die Zentralbanken divergierenden Inflationsraten mit Hilfe von Stützungskäufen der schwachen bzw. Verkäufen der starken Währung begegnen, kann der Zusammenhang zwischen der Preis- und der Wechselkursentwicklung nicht so direkt sein wie zwischen frei floatenden Währungen. Das wirft die Frage auf, ob sich systematische Zentralbankeingriffe, die in der Regel Auf- bzw. Abwertungstendenzen entgegengerichtet sein werden, in typischen Verlaufsmustern der Wechselkurse niederschlagen.

Zum anderen kommt dem Element der *Spekulation* in einem solchen Wechselkurssystem, das Realignments ausdrücklich vorsieht, besondere Bedeutung zu. Die beteiligten Zentralbanken haben sich ja aus währungspolitischen Gründen verpflichtet, solange bestimmte An- und Verkaufskurse zu garantieren, bis neue Leitkurse festgelegt werden. Ein Marktteilnehmer, der solch ein Realignment vorhersieht, kann die durch diese Garantieerklärung bedingte Starrheit

[52] Diese Arbeit kursierte bereits seit 1989 als Arbeitspapier, so daß die ersten Modifikationen und empirischen Untersuchungen schon vor 1991 erschienen.

[53] Vgl. z.B. BERTOLA und CABALLERO [1991] oder BERTOLA und SVENSSON [1993].

der Kurse ausnutzen, indem er Kredite in einer unter Abwertungsverdacht stehenden Währung aufnimmt und damit eine unter Aufwertungsverdacht stehende Währung bei den Zentralbanken kauft, bis die Zentralbanken nicht mehr willens oder in der Lage sind, die starke Währung zu dem garantierten Kurs zur Verfügung zu stellen. Dies führt dann sofort zu einem deutlichen Kursanstieg der stärkeren Währung gegenüber der schwächeren, den die Zentralbanken dann durch ein Realignment bestätigen werden (falls nicht sogar ein Land oder mehrere Länder aus dem Wechselkursverbund ausscheren). Der Spekulant kann diesen Kursanstieg ausnutzen, seine Guthaben in der starken Währung zu einem höheren Kurs verkaufen, als er sie erworben hat, und behält nach Tilgung seiner Kredite in der schwachen Währung einen Gewinn.

Marktteilnehmer, die über Devisenbestände (oder Kredite) in der Größenordnung der Devisenreserven und Finanzierungsfaszilitäten der beteiligten Zentralbanken verfügten, stellten eine ernsthafte Bedrohung für das EWS dar. Der bekannteste Fall einer solchen Spekulation war der gegen das britische Pfund und die italienische Lira im Jahre 1992 (unter anderem durch den Amerikaner George Soros), der dazu führte, daß diese beiden Länder am 17. September 1992 aus dem EWS austraten, nachdem sie praktisch die gesamten nationalen Devisenbestände zur Verteidigung der schließlich nicht haltbaren Leitkurse eingesetzt hatten.

Spekulative Krisen des Währungssystems waren regelmäßig mit stark schwankenden Wechselkursen verbunden, die die der Marktentwicklung entgegengerichteten Eingriffe der Zentralbanken und natürlich die „Spekulation gegen die Spekulanten" hervorrief. Kurzfristig stark fluktuierende Wechselkurse (d.h. hohe Volatilität) in der Nähe einer Interventionsgrenze können also ein Anzeichen für eine solche spekulative Phase darstellen, auch wenn sie – ex post – tatsächlich *nicht* zu einem Realignment führten. Eine weitere Begleiterscheinung solcher Perioden ist die in der UIP zum Ausdruck gebrachte Risikoprämie in Form von höheren Zinsen, die die Marktteilnehmer für die Anlage in einer unter Abwertungsverdacht stehenden Währung verlangen. Es ist deshalb verschiedentlich vorgeschlagen worden, hohe Zinsdifferenzen als Indikator für eine spekulative Phase zu verwenden.[55]

Wie in Kapitel 3.1.2.2 erläutert wurde, war der Interventionsmechanismus des EWS nicht an die ECU-Leitkurse, sondern an das bilaterale Paritätengitter gekoppelt. Ferner wurde in Kapitel 3.1.2.4 erläutert, daß die DM faktisch die Rolle einer Ankerwährung des EWS einnahm. Deswegen ist es sinnvoll und üblich, als Grundlage empirischer Untersuchungen die bilateralen Wechselkurse zwischen der DM und den übrigen nationalen Währungen zu ver-

[54] Vgl. z.B. SMITH und SPENCER [1991] oder DE JONG [1994].

[55] Vgl. z.B. BEKAERT und GRAY [1998].

wenden (und nicht etwa die ECU-Kurse, die in offiziellen Publikationen wie z.B. den Bundesbankberichten häufig dargestellt werden).

3.3 Die empirische Untersuchung von Wechselkursen im EWS mit Hilfe von Regimewechselmodellen

In der empirischen Literatur über das EWS sind verschiedene Versuche unternommen worden, das wechselnde spekulative und nicht-spekulative Verhalten der Marktteilnehmer mit Hilfe von Regimewechselmodellen zu beschreiben.[56] ENGEL und HAKKIO [1996] schätzen inhomogene Markov-Modelle, bei denen der Regimeindikator von der relativen Position im Wechselkursband in der Vorperiode abhängt, für den Kurs der DM gegen den Französischen Franc und die Italienische Lira mit Daten von 1979 bis 1992 bzw. 1993. BEKAERT und GRAY [1998] verwenden für ein Regimewechselmodell für den DM/Franc-Kurs im selben Zeitraum die Zinsdifferenz zwischen den beiden Ländern als Regimeindikator. Der jeweils verwendete Regimewechselmechanismus wird aber in der Regel kaum durch Tests statistisch abgesichert, die sicherstellen, daß mehrere Zustände vorliegen und daß sie auf die angegebene Weise zustande kommen. Für dieses Defizit an statistischer Untermauerung des Mechanismus der Parametervariation gilt immer noch die Forderung von Boothe und Glassman:

> *„Further research on the hypothesis of time-varying parameters should focus on specifying and testing models of systematic parameter variation over time."*[57]

Außerdem konzentrieren sich die Untersuchungen auf die frühe Phase des EWS bis zur Krise 1992/93. Es stellt sich die Frage, ob die EWS-Kurse auch danach, insbesondere nach der Erweiterung der Bandbreiten, Anzeichen von Regimewechseln zeigen, die sich durch spekulative Phasen erklären lassen, und wie sich gegebenenfalls die Kurse in den Zuständen verhielten.

Im empirischen Teil dieser Arbeit soll daher zum einen gezeigt werden, wie spekulative Phasen in der jüngeren Geschichte des EWS mit Hilfe von Regimewechselmodellen beschrieben und erklärt werden können. Hierfür ist es notwendig, relativ kurzfristige Daten zu untersuchen (hier werden Wochenwerte verwendet), für die keine entsprechenden makroökonomischen Werte zur Verfügung stehen. Deswegen stehen in dieser Betrachtung statistische und zeitreihentheoretische Aspekte im Vordergrund.

---

[56] Diese Arbeiten schließen an frühere Untersuchungen über Regimewechselmodelle für frei floatende Wechselkurse an, vgl. etwa ENGLE und HAMILTON [1990], LEE [1991] und KAEHLER und MARNET [1994].

[57] BOOTHE und GLASSMAN [1987], S. 316.

Zum anderen soll die PPP im Kontext von Regimewechselmodellen untersucht werden. In den letzten Jahren ist die absolute PPP hauptsächlich mit Hilfe von Kointegrationsanalysen überprüft worden, da es Hinweise darauf gibt, daß die Niveaus der betroffenen Größen nichtstationäre Prozesse darstellen.[58] Da aber für Regimewechselmodelle, so wie sie in dieser Arbeit behandelt werden, die Konsistenz der Schätzer i.a. nur für stationäre Zeitreihen gesichert ist (vgl. dazu Kapitel 6.1), wird hier die relative Form der PPP untersucht. Allgemein wird davon ausgegangen, daß sich die relative PPP für kurzfristige Werte empirisch nicht bestätigen läßt, und daß ein Wechselkursmechanismus wie das EWS einen derartigen Zusammenhang gar nicht zuläßt. Diese Annahme soll mit Regimewechselmodellen, die Verhaltensänderungen ausdrücklich berücksichtigen, anhand von Monatsdaten überprüft werden.

Beide Untersuchungen werden an zwei Wechselkursen im EWS, dem zwischen dem Französischen Franc (FF) und der DM sowie dem des Niederländischen Gulden (NG) und der DM, durchgeführt. Die DM wurde in beiden Fällen als Bezugsgröße gewählt, da sie (wie oben beschrieben) faktisch die Rolle einer Ankerwährung im EWS übernommen hatte. Der Franc als wichtigste Währung im EWS neben der DM war starken Schwankungen unterworfen und repräsentiert damit eine Währung, die anfällig für Spekulation war. Im Gegensatz dazu stellte der Gulden mit seiner starken Bindung an die DM eine Währung mit geringen Kursausschlägen dar, für die Spekulation – und damit vermutlich auch Verhaltensänderungen – von geringerer Bedeutung gewesen sein dürfte.[59]

Von diesen insgesamt vier Untersuchungen wird die erste (basierend auf Wochendaten für den Wechselkurs zwischen Franc und DM) verwendet, um die wichtigsten Aspekte der Modellierung von Regimewechselmodellen – Schätzung, Tests und Prognosen – direkt im Anschluß an die theoretische Darstellung in den entsprechenden Kapiteln ausführlich zu veranschaulichen. Die restlichen drei Untersuchungen werden in etwas knapperer Form in Kapitel 10 präsentiert.

[58] Vgl. z.B. MARK [1990], CHEUNG und LAI [1993] oder DUEKER [1993].

[59] Der empirische Teil der Arbeit ist ähnlich aufgebaut wie der in DIEBOLD [1988], in dem verschiedene floatende Wechselkurse im Zusammenhang von (G)ARCH-Modellen an Wochendaten und im Kontext von PPP und Integration/Kointegration anhand von Monatsdaten untersucht werden.

4 ML-Schätzung

Für die Schätzung von Regimewechselmodellen hat sich das Maximum-Likelihood-(ML-)Verfahren durchgesetzt. Dafür gibt es verschiedene Gründe. Zum einen bieten sich in den meisten Fällen keine anderen Schätzer unmittelbar an (mit Ausnahme der Varianzschätzer, worauf an den entsprechenden Stellen genauer eingegangen wird). Zum anderen ist die ML-Schätzung eng mit den asymptotischen Testverfahren verknüpft, die im Zusammenhang von Regimewechselmodellen häufig angewendet werden.[60]

Die ML-Schätzer sind – je nach Modelltyp – denen für klassische Regressions- bzw. Zeitreihenmodelle (aufgrund der Verwandtschaft der Regimewechselmodelle zu ihnen) sehr ähnlich. Für das klassische Regressionsmodell ist der ML-Schätzer des Regressionsparameters gleich dem OLS-(*Ordinary Least Squares*-)Schätzer, der ML-Schätzer der Störvarianz unterscheidet sich nur um einen nicht von den Beobachtungen abhängigen Faktor von dem normalerweise verwendeten erwartungstreuen Schätzer, und diese beiden sind asymptotisch gleich. Ähnliches gilt für AR-Modelle. Deswegen sind die ML-Schätzer für homoskedastische Regimewechselmodelle den OLS-Schätzern gewöhnlicher linearer Modelle ähnlich, und für ihre Berechnung kann auf das OLS-Verfahren zurückgegriffen werden. Ähnliche Analogien bestehen zwischen den ML-Schätzern für heteroskedastische Regimewechselmodelle und den GLS-(*Generalised Least Squares*-)Schätzern heteroskedastischer linearer Modelle Dieser Aspekt ist besonders für die praktische Arbeit mit Regimewechselmodellen von Bedeutung und wird deshalb im folgenden besonders beachtet.

Die ML-Schätzer für Modelle mit latenten Zuständen (d.h. Mischungs- und Markov-Modelle) können ihrerseits als eine Verallgemeinerung derjenigen für Modelle mit beobachtbaren Zuständen (also insbesondere Schwellenmodelle mit bekanntem Schwellenwert) angesehen werden. Um diesen Zusammenhang deutlich zu machen und um die ML-Schätzer für Mischungs- und Markov-Modelle übersichtlich darstellen zu können, werden im folgenden zuerst Schwellenmodelle ausführlich behandelt.

4.1 Stochastische Spezifikation

In diesem Abschnitt werden die allgemeinen Voraussetzungen für Regimewechselmodelle mit zwei Zuständen spezifiziert und einige Bezeichnungen eingeführt. Diese werden dann in den folgenden Abschnitten für die einzelnen Modelltypen präzisiert.

[60] Vg. hierzu Kapitel 6.

Die endogene Variable sei

$$\mathbf{y} = (y_1,\ldots,y_T)',$$

die Regressoren (die vorherbestimmten Größen) seien

$$\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_K)$$

und $$\mathbf{W} = (\mathbf{w}_1,\ldots,\mathbf{w}_L)\ ,$$

wobei $$\mathbf{x}_i = (x_{1,i},\ldots,x_{T,i})' \text{ für } i = 1,\ldots,K$$

und $$\mathbf{w}_i = (w_{1,i},\ldots,w_{T,i})' \text{ für } i = 1,\ldots,L$$

Spaltenvektoren mit den Werten eines festen Regressors zu verschiedenen Zeitpunkten darstellen. **X** und **W** enthalten keine gemeinsamen Spalten.

Die Werte der vorherbestimmten Größen zum Zeitpunkt t werden mit den Zeilenvektoren

$$\mathbf{X}_t = (x_{t,1},\ldots,x_{t,K})$$

bzw. $$\mathbf{W}_t = (w_{t,1},\ldots,w_{t,L})$$

bezeichnet.

Ein Teil der Parameter des Modells zum Zeitpunkt t hängt von dem jeweiligen Zustand ab, der durch die Zustandsvariable s_t beschrieben wird, die die zwei möglichen Werte 1 und 2 annehmen kann. Die Zustandsvariablen bilden den Zustandsvektor

$$\mathbf{s} = (s_1,\ldots,s_T)'\ .$$

Falls **X** oder **W** (oder beide) verzögert endogene Größen der maximalen Laglänge m enthalten, so sei

$$\mathbf{y} = (y_{-m+1},\ldots,y_0,y_1,\ldots,y_T)',$$

$$\mathbf{x}_i = (x_{-m+1,i},\ldots,x_{0,i},x_{1,i},\ldots,x_{T,i})' \text{ für } i = 1,\ldots,K$$

und $$\mathbf{w}_i = (w_{-m+1,i},\ldots,w_{0,i},w_{1,i},\ldots,w_{T,i})' \text{ für } i = 1,\ldots,L\ .$$

Die übrigen Regressoren (die exogenen Größen) sind entweder nichtstochastisch, oder stochastisch und unabhängig von dem Störprozeß

$$\mathbf{u} = (u_1,\ldots,u_T)'\ .$$

Für die Verteilung der Störvariablen gilt

$$\mathbf{u} \sim \mathrm{N}(0,\Omega),\ \Omega = \begin{pmatrix} \sigma_{s_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{s_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{s_T}^2 \end{pmatrix},\ \text{wobei } \sigma_1{}^2 > 0 \text{ und } \sigma_2{}^2 > 0\ .$$

Die Störvariablen zu verschiedenen Zeitpunkten sind also unabhängig, und es wird Heteroskedastie in dem Sinne betrachtet, daß die Störvarianz von dem jeweiligen Zustand abhängen kann.

Ein Regimewechselmodell lautet dann allgemein

$$y_t = \begin{cases} \mathbf{W}_t\alpha + \mathbf{X}_t\beta_1 + u_t \text{, falls } s_t = 1 \\ \mathbf{W}_t\alpha + \mathbf{X}_t\beta_2 + u_t \text{, falls } s_t = 2 \text{,} \end{cases}$$

wobei

$$\beta_1 = (\beta_{1,1},\ldots,\beta_{1,K})', \ \beta_2 = (\beta_{2,1},\ldots,\beta_{2,K})' \text{ und } \alpha = (\alpha_1,\ldots,\alpha_L)' \ .$$

Es sind demnach nur die Regressionsparameter von **X**, nicht aber die von **W**, vom jeweiligen Zustand abhängig.

Dieses allgemeine Modell enthält u.a. folgende *Spezialfälle*:

(1) $\mathbf{X} = \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Dies ist ein lineares Regressionsmodell mit einer speziellen Form von Heteroskedastie.

(2) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} = \mathbf{0}$, $\sigma_1^2 = \sigma_2^2$: Alle Regressionsparameter sind zustandsabhängig, es liegt Homoskedastie vor.

(3) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} = \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Alle Regressionsparameter sind zustandsabhängig, das Modell ist heteroskedastisch.

(4) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} \neq \mathbf{0}$, $\sigma_1^2 = \sigma_2^2$: Es gibt zustandsabhängige und -unabhängige Regressionsparameter, es liegt Homoskedastie vor.

(5) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} \neq \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Es gibt zustandsabhängige und -unabhängige Regressionsparameter, das Modell ist heteroskedastisch.

Ein interessanter Fall, der durch diese Spezifikationen nicht erfaßt wird, liegt vor, wenn weitere Regressoren $\mathbf{V}^1$ die abhängige Variable nur in Zustand 1 beeinflussen und andere Regressoren $\mathbf{V}^2$ nur in Zustand 2. Der allgemeinste Fall wäre dann also

$$y_t = \begin{cases} \mathbf{W}_t\alpha + \mathbf{X}_t\beta_1 + \mathbf{V}_t^1\gamma_1 + u_t \text{, falls } s_t = 1 \\ \mathbf{W}_t\alpha + \mathbf{X}_t\beta_2 + \mathbf{V}_t^2\gamma_2 + u_t \text{, falls } s_t = 2 \end{cases} .$$

Um die Darstellung übersichtlich zu halten, wurde dieser Fall nicht gesondert in die obige Auflistung aufgenommen. Er läßt sich aber ohne weiteres mit den weiter unten dargestellten Methoden behandeln. In den folgenden Kapiteln wird an den entsprechenden Stellen darauf hingewiesen, wie sich die dargestellten Ergebnisse auf diesen Fall verallgemeinern lassen.

4.2 Schwellenmodelle

Bei diesen Modellen hängt die Zustandsvariable direkt in Form einer Treppenfunktion von einem Zustandsindikator z_t ab:

$$s_t = \begin{cases} 1 \text{ falls } z_t < c \\ 2 \text{ falls } z_t \geq c\,. \end{cases}$$

Hierbei ist z_t ebenso wie die exogenen Größen entweder nichtstochastisch, oder stochastisch und unabhängig von dem Störprozeß $\mathbf{u}$.[61]

Es wird zuerst der Fall betrachtet, bei dem $\mathbf{X}$ und $\mathbf{W}$ keine verzögert endogenen Variablen enthalten. Falls einige oder alle exogenen Größen stochastisch sind, wird durch das Modell nur die bedingte Verteilung von $\mathbf{y}$ gegeben diese exogenen Größen spezifiziert. In diesem Fall sind die im folgenden betrachteten Wahrscheinlichkeitsdichten immer als "gegeben $\mathbf{X}$, $\mathbf{W}$ und $\mathbf{z}$" zu verstehen, was aber um einer besseren Übersichtlichkeit willen nicht ausdrücklich notiert wird. Es gilt

$$\begin{aligned} f(y_t) &= f(y_t, s_t = 1) + f(y_t, s_t = 2) \\ &= \begin{cases} f(u_t = y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1), \text{falls } z_t < c \\ f(u_t = y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2), \text{falls } z_t \geq c \end{cases} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)^2}{2\sigma_1^2}}, \text{falls } z_t < c \\ \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2)^2}{2\sigma_2^2}}, \text{falls } z_t \geq c\,. \end{cases} \end{aligned}$$

Da die Fehlervariablen u_t zu verschiedenen Zeitpunkten unabhängig (und unabhängig von den Regressoren) sind, gilt

$$\begin{aligned} f(\mathbf{y}) &= \prod_{t=1}^{T} f(y_t) \\ &= \prod_{z_t < c} f(y_t) \prod_{z_t \geq c} f(y_t) \\ &= \prod_{z_t < c} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)^2}{2\sigma_1^2}} \prod_{z_t \geq c} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2)^2}{2\sigma_2^2}}\,. \end{aligned}$$

Wird die Anzahl der z_t mit $z_t < c$ (bzw. $z_t \geq c$) mit T_1 (bzw. T_2) bezeichnet, so gilt

[61] Die Ausführungen im diesem Kapitel sind für Schwellenmodelle formuliert. Sie sind jedoch auch für andere Modelle mit beobachtbarem Zustandsindikator gültig (es müssen lediglich die Ausdrücke "$z < c$" und "$z \geq c$" durch die entsprechenden Bedingungen für das Vorliegen der einzelnen Zustände ersetzt werden).

$$\begin{aligned}\log(f(\mathbf{y})) &= \sum_{t=1}^{T}\log(f(y_t)) \\ &= -\frac{T}{2}\log(2\pi) - T_1\log(\sigma_1) - T_2\log(\sigma_2) \\ &\quad - \sum_{z_t<c}\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)^2}{2\sigma_1^2} - \sum_{z_t\geq c}\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2)^2}{2\sigma_2^2}.\end{aligned}$$

Falls **X** oder **W** (oder beide) verzögert endogene Größen der maximalen Laglänge m enthalten, so gilt

$$f(\mathbf{y}) = f(y_T,\ldots,y_{-m+1}) = \left(\prod_{t=1}^{T} f(y_t \mid y_{t-1},\ldots,y_{t-m})\right) f(y_0,\ldots,y_{-m+1}),$$

also

$$f(\mathbf{y}|y_0,\ldots,y_{-m+1}) = \prod_{t=1}^{T} f(y_t|y_{t-1},\ldots,y_{t-m}),$$

und

$$f(y_t \mid y_{t-1},\ldots,y_{t-m}) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_t-\mathbf{W}_t\alpha-\mathbf{X}_t\beta_1)^2}{2\sigma_1^2}}, & \text{falls } z_t < c \\ \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y_t-\mathbf{W}_t\alpha-\mathbf{X}_t\beta_2)^2}{2\sigma_2^2}}, & \text{falls } z_t \geq c. \end{cases}$$

Der obige Ausdruck für $\log(f(\mathbf{y}))$ stellt also, falls verzögert endogene Variablen unter den Regressoren sind, die bedingte Log-Likelihood gegeben $y_0,\ldots,y_{-m+1}$ (sowie **X**, **W** und **z**) dar. Dies wird im folgenden der Übersichtlichkeit halber in der Notation nicht mehr ausdrücklich berücksichtigt. Die weiter unten berechneten ML-Schätzer sind hier (auf die ersten m Werte der endogenen Variablen) bedingte ML-Schätzer.[62] Berücksichtigt man diesen Unterschied in der Interpretation der Wahrscheinlichkeitsdichten, behalten alle folgenden Ausführungen auch für dynamische und reine Zeitreihenmodelle ihre Gültigkeit. Liegen noch Regressoren $\mathbf{V}^1$ in Zustand 1 und $\mathbf{V}^2$ in Zustand 2 vor, so ist in den obigen Ausdrücken $(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_i)$ durch $(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_i - \mathbf{V}^i\gamma_i)$ für $i = 1, 2$ zu ersetzen.

Die ML-Schätzer werden zuerst (in Kapitel 4.2.1) für Schwellenmodelle mit bekanntem Schwellenwert hergeleitet, also für Modelle, bei denen die Abfolge der Zustände von vornherein bekannt ist. Liefert der Zustandsindikator nur unvollständige Informationen über die vorliegenden Zustände (etwa im Falle eines Schwellenmodells mit unbekanntem Schwellen-

[62] Diese bedingten ML-Schätzer sind für lineare Modelle asymptotisch äquivalent zu den exakten ML-Schätzern, vgl. JUDGE u.a. [1985], S. 237 f., oder HAMILTON [1994b], Kap. 5.3.

wert oder eines verallgemeinerten Schwellenmodells), so ergeben sich besondere Probleme bei der ML-Schätzung, die in Kapitel 4.2.2 behandelt werden.

4.2.1 Bekannter Schwellenwert

Falls der Schwellenwert bekannt ist, also das jeweilige Regime beobachtbar ist, lassen sich in den meisten Fällen aus den Normalgleichungen, die in Kapitel 4.2.1.1 berechnet werden, geschlossene Ausdrücke für die ML-Schätzer herleiten, die in Kapitel 4.2.1.2 angegeben sind. Die übrigen Fälle, für die sich die Normalgleichungen nicht auflösen lassen, werden in Kapitel 4.2.1.3 behandelt.

4.2.1.1 Bestimmung der Normalgleichungen

Durch Differentiation nach den Parametern erhält man folgende Normalgleichungen:

$$\frac{\partial \log(f(\mathbf{y}))}{\partial \beta_1} = \mathbf{0} \quad \Leftrightarrow$$

$$\sum_{z_t<c} \mathbf{X}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1) = \mathbf{0} \ , \tag{a1}$$

$$\frac{\partial \log(f(\mathbf{y}))}{\partial \beta_2} = \mathbf{0} \quad \Leftrightarrow$$

$$\sum_{z_t\geq c} \mathbf{X}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2) = \mathbf{0} \ , \tag{a2}$$

$$\frac{\partial \log(f(\mathbf{y}))}{\partial \alpha} = \mathbf{0} \quad \Leftrightarrow$$

$$\sum_{z_t<c} \frac{\mathbf{W}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)}{2\sigma_1^2} + \sum_{z_t\geq c} \frac{\mathbf{W}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2)}{2\sigma_2^2} = \mathbf{0} \ . \tag{b}$$

Ist $\sigma_1^2 = \sigma_2^2$, so lautet diese Gleichung

$$\sum_{z_t<c} \mathbf{W}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1) + \sum_{z_t\geq c} \mathbf{W}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2) = \mathbf{0} \ . \tag{b'}$$

$$\frac{\partial \log(f(\mathbf{y}))}{\partial \sigma_1} = 0 \quad \Leftrightarrow$$

$$\sigma_2^2 = \frac{\sum_{z_t<c}(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)^2}{T_1} \ , \tag{c1}$$

$$\frac{\partial \log(f(\mathbf{y}))}{\partial \sigma_2} = 0 \quad \Leftrightarrow$$

$$\sigma_2^2 = \frac{\sum_{z_t\geq c}(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2)^2}{T_2} \ . \tag{c2}$$

Ist $\sigma_1^2 = \sigma_2^2$, so erhält man anstelle von (c1) und (c2) die Gleichung

$$\sigma^2 = \frac{\sum_{z_t<c}(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)^2 + \sum_{z_t\geq c}(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_2)^2}{T}. \qquad \text{(c)}$$

4.2.1.2 Berechnung von Parameterschätzern durch Hilfsregressionen

In diesem Abschnitt wird erläutert, wie die (bedingten) ML-Schätzer, die sich aus den Normalgleichungen des letzten Abschnitts ergeben, mit Hilfe von Einfachregressionen berechnet werden können.

(1) $\mathbf{X} = \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Hier läßt sich kein geschlossener Ausdruck für die ML-Schätzer angeben. Dieser Fall wird in Kapitel 4.2.1.3 behandelt.

(2) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} = \mathbf{0}$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$: Die Normalgleichungen sind (a1), (a2) und (c). Die Gleichungen (a1) und (a2) sind die Gleichungen separater Regressionen von $\mathbf{y}$ auf $\mathbf{X}$ für die Zeiträume, die durch $z_t < c$ bzw. $z_t \geq c$ gekennzeichnet sind. Diese Regressionen kann man auch folgendermaßen berechnen: Sei $\mathbf{X}^1$ die Matrix, deren t-te Zeile gleich $\mathbf{X}_t$ ist, falls $s_t = 1$, und 0 sonst. Analog sei $\mathbf{X}^2$ definiert. Es gilt also $\mathbf{X} = \mathbf{X}^1 + \mathbf{X}^2$. Sei

$$\overline{\mathbf{X}} = \begin{pmatrix}\mathbf{X}^1 & \mathbf{X}^2\end{pmatrix} \text{ und } \beta = \begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix}.$$

Dann sind die Normalgleichungen für die Parameter des gewöhnlichen linearen Modells

$$\mathbf{y} = \overline{\mathbf{X}}\beta + \mathbf{u}\,,\ \mathbf{u} \sim \mathrm{N}(0, \sigma^2\mathbf{I})$$

dieselben wie die in Kapitel 4.2.1.1 hergeleiteten. Sie lauten nämlich in Matrixform

$$\overline{\mathbf{X}}'(\mathbf{y} - \overline{\mathbf{X}}\beta) = \mathbf{0}$$

$$\Leftrightarrow \begin{pmatrix}(\mathbf{X}^1)' \\ (\mathbf{X}^2)'\end{pmatrix}\left(\mathbf{y} - \begin{pmatrix}\mathbf{X}^1\beta_1 & \mathbf{X}^2\beta_2\end{pmatrix}\right) = \mathbf{0}$$

(dies sind die Gleichungen (a1) und (a2), da nach Konstruktion $(\mathbf{X}^1)'\mathbf{X}^2 = (\mathbf{X}^2)'\mathbf{X}^1 = \mathbf{0}$), sowie

$$\hat{\sigma}^2 = \frac{1}{T}(\mathbf{y} - \overline{\mathbf{X}}\hat{\beta})'(\mathbf{y} - \overline{\mathbf{X}}\hat{\beta})$$

(dies ist Gleichung (c)). Also sind $\hat{\beta}_1$ und $\hat{\beta}_2$ die OLS-Schätzer einer Regression von $\mathbf{y}$ auf die Spalten von $\mathbf{X}^1$ und $\mathbf{X}^2$, d.h.

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \hat{\beta} = (\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\mathbf{y} \ ,^{63}$$

und $\hat{\sigma}^2$ der ML-Schätzer der Fehlervarianz dieses Modells (ohne Korrektur um die Freiheitsgrade).

(3) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} = \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Die Normalgleichungen für β_1 und β_2 sind wieder (a1) und (a2). Also sind $\hat{\beta}_1$ und $\hat{\beta}_2$ dieselben Schätzer wie bei **(2)**. Obwohl in diesem Fall Heteroskedastie (mit unbekannter Varianz-Kovarianz-Matrix) vorliegt, liefert die OLS-Schätzung wegen der speziellen Form von Heteroskedastie also auch ML-Schätzer für β_1 und β_2 (was im allgemeinen nicht der Fall ist, vgl. Kapitel 4.2.1.3).

Die Normalgleichungen für σ_1^2 und σ_2^2 sind (c1) und (c2), die ML-Schätzer sind jeweils das arithmetische Mittel der geschätzten Fehler der Hilfsregression in den Zeiträumen, die durch $z_t < c$ bzw. $z_t \geq c$ gekennzeichnet sind. Definiert man $\mathbf{y}^1$ und $\mathbf{y}^2$ analog zu $\mathbf{X}^1$ und $\mathbf{X}^2$ (d.h. $\mathbf{y}_t^1 = y_t$, falls $s_t = 1$, und $\mathbf{y}_t^1 = 0$ sonst, und entsprechend für $\mathbf{y}^2$, also $\mathbf{y} = \mathbf{y}^1 + \mathbf{y}^2$), so gilt

$$\hat{\sigma}_i^2 = \frac{1}{T_i}(\mathbf{y}^i - \mathbf{X}^i\hat{\beta}_i)'(\mathbf{y}^i - \mathbf{X}^i\hat{\beta}_i) \text{ für } i = 1, 2.$$

Der Schätzer $\hat{\sigma}^2$ aus **(2)** ist das mit den relativen Häufigkeiten der beiden Zustände gewichtete Mittel von $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$, d.h.

$$\hat{\sigma}^2 = \frac{T_1}{T}\hat{\sigma}_1^2 + \frac{T_2}{T}\hat{\sigma}_2^2 \ .$$

(4) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} \neq \mathbf{0}$, $\sigma_1^2 = \sigma_2^2$: Die Normalgleichungen sind (a1), (a2), (b') und (c). Da (a1), (a2) und (b') die Parameter α, β_1 und β_2 enthalten, kann man die ML-Schätzer jetzt nicht mehr mit separaten Regressionen berechnen. Statt dessen bietet sich folgende Hilfsregression an: Seien $\mathbf{X}^1$ und $\mathbf{X}^2$ wie bei **(2)** definiert. Sei jetzt

$$\overline{\mathbf{X}} = \begin{pmatrix} \mathbf{X}^1 & \mathbf{X}^2 & \mathbf{W} \end{pmatrix} \text{ und } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \alpha \end{pmatrix} .$$

Dann sind wieder die Normalgleichungen für die Parameter des Modells

$$\mathbf{y} = \overline{\mathbf{X}}\beta + \mathbf{u} \ , \ \mathbf{u} \sim \mathrm{N}(0, \sigma^2\mathbf{I})$$

dieselben wie die in Kapitel 4.2.1.1 hergeleiteten. Sie lauten hier für β

[63] Hier muß man natürlich die Invertierbarkeit der Momentenmatrix voraussetzen. Dies gilt analog auch für die folgenden Fälle.

$$\overline{\mathbf{X}}'(\mathbf{y}-\overline{\mathbf{X}}\beta)=\mathbf{0}$$

$$\Leftrightarrow \begin{pmatrix} (\mathbf{X}^1)' \\ (\mathbf{X}^2)' \\ \mathbf{W} \end{pmatrix} \left(\mathbf{y}-\left(\mathbf{X}^1\beta_1 \quad \mathbf{X}^2\beta_2 \quad \mathbf{W}\alpha \quad \right)\right)=\mathbf{0}$$

(dies sind die Gleichungen (a1), (a2) und (b')), sowie für σ^2

$$\hat{\sigma}^2=\frac{1}{T}(\mathbf{y}-\overline{\mathbf{X}}\hat{\beta})'(\mathbf{y}-\overline{\mathbf{X}}\hat{\beta}) \ .$$

(das ist Gleichung (c)). Also sind $\hat{\beta}_1$ und $\hat{\beta}_2$ die OLS-Schätzer einer Regression von **y** auf die Spalten von $\mathbf{X}^1$, $\mathbf{X}^2$ und **W**, und $\hat{\sigma}^2$ das arithmetische Mittel der geschätzten Fehler dieser Regression.

(5) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} \neq \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Die Normalgleichungen sind (a1), (a2), (b), (c1) und (c2). Da (b) alle Parameter enthält, aber nicht linear ist, läßt sich hier kein geschlossener Ausdruck für die ML-Schätzer angeben. Sie sind die ML-Schätzer des Modells

$$\mathbf{y}=\overline{\mathbf{X}}\beta+\mathbf{u}\,,\ \mathbf{u}\sim \mathrm{N}(0,\Omega),\ \Omega=\begin{pmatrix} \sigma_{s_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{s_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{s_T}^2 \end{pmatrix},$$

mit $\overline{\mathbf{X}}$ und β wie in **(4)**. Dieser Fall wird in Kapitel 4.2.1.3 behandelt.

Der maximale Wert der Log-Likelihood-Funktion, der für die später zu besprechenden asymptotischen Tests benötigt wird, läßt sich leicht berechnen als Wert der konzentrierten Log-Likelihood-Funktion an der Stelle $(\hat{\sigma}_1,\hat{\sigma}_2)$ bzw. $\hat{\sigma}$:

$$\max(\log(f(\mathbf{y}))=\begin{cases} -\frac{T}{2}(\log(2\pi)+1)-\frac{T_1}{2}\log\hat{\sigma}_1^2-\frac{T_2}{2}\log\hat{\sigma}_2^2\,, & \text{falls } \sigma_1^2\neq\sigma_2^2 \\ -\frac{T}{2}(\log(2\pi)+1+\log\hat{\sigma}^2)\,, & \text{falls } \sigma_1^2=\sigma_2^2\,. \end{cases}$$

Liegen noch Regressoren $\mathbf{V}^1$ in Zustand 1 und $\mathbf{V}^2$ in Zustand 2 vor, so sind diese Aussagen weiter gültig, wenn man in den obigen Ausdrücken $\mathbf{X}^i$ durch $\left(\mathbf{X}^i \quad \mathbf{V}^i\right)$ und β_i' durch $(\beta_i'\gamma_i')$ für $i=1, 2$ ersetzt, also z.B. in Fall **(5)**

$$\overline{\mathbf{X}}=\left(\mathbf{X}^1 \quad \mathbf{V}^1 \quad \mathbf{X}^2 \quad \mathbf{V}^2 \quad \mathbf{W}\right) \text{ und } \beta=\left(\beta_1' \quad \gamma_1' \quad \beta_2' \quad \gamma_2' \quad \alpha'\right)' \ .$$

4.2.1.3 Berechnung der ML-Schätzer für heteroskedastische Modelle mit regimeunabhängigen Parametern

Für die praktische Berechnung der Schätzer in den verbleibenden Fällen, also **(1)** und **(5)**, muß man auf iterative Verfahren zurückgreifen. Die ML-Schätzer für diese Modelle lassen sich im Prinzip mit den üblichen Verfahren zur Bestimmung (lokaler) Maxima nichtlinearer Funktionen bestimmen, vgl. hierzu Kapitel 4.3.1. Eine rechnerisch sehr einfache Alternative zu diesen Methoden ist eine Variante des *Oberhofer-Kmenta-Verfahrens*[64], das im folgenden für die vorliegende Situation kurz dargestellt und erläutert wird.[65]

In Kapitel 4.2.1.2 wurde gezeigt, daß sich in beiden hier betrachteten Fällen das Modell in der Form

$$\mathbf{y} = \mathbf{Z}\vartheta + \mathbf{u}\,,\ \mathbf{u} \sim \mathrm{N}(0, \Omega),\ \Omega = \begin{pmatrix} \sigma_{s_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{s_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{s_T}^2 \end{pmatrix}$$

darstellen läßt (bei **(1)** mit $\mathbf{Z} = \mathbf{W}$ und $\vartheta = \alpha$, und bei **(5)** mit $\mathbf{Z} = \overline{\mathbf{X}}$ und $\vartheta = (\beta_1', \quad \beta_2', \quad \alpha')'$). Die folgenden Ausführungen gelten also sinngemäß für beide Fälle.

Dieses Modell ist ein verallgemeinertes Regressionsmodell mit unbekannter Varianz-Kovarianz-Matrix der Störvariablen, deren parametrische Form jedoch bekannt ist. Das Oberhofer-Kmenta-Verfahren besteht darin, abwechselnd

- die unbekannten Varianzen mit Hilfe der letzten geschätzten Regressionsparameter aus den geschätzten Fehlern für die jeweiligen Zustände zu berechnen und
- die Regressionsparameter mit einer FGLS-(*Feasible Generalised Least Squares*-) Schätzung unter Verwendung der letzten geschätzten Varianzen zu berechnen.

Eine FGLS-Schätzung ist eine GLS-(*Generalised Least Squares*-)Schätzung mit geeignet geschätzter Varianz-Kovarianz-Matrix der Störvariablen, d.h. allgemein gilt

$$\hat{\vartheta} = (\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{y}\ .$$

Im vorliegenden Fall ist

[64] Vgl. hierzu OBERHOFER und KMENTA [1974], MAGNUS [1978] und GREENE [1997], S. 513 ff.

[65] Das Oberhofer-Kmenta-Verfahren ist wesentlich allgemeiner als hier dargestellt und läßt sich auch auf die ML-Schätzung anderer Formen von Heteroskedastie anwenden. Für autokorrelierte Fehlerprozesse führt es z.B. auf das *Cochrane-Orcutt-Verfahren*, vgl. COCHRANE und ORCUTT [1949].

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_{s_1}^2 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_{s_2}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \hat{\sigma}_{s_T}^2 \end{pmatrix},$$

und obiger FGLS-Schätzer berechnet sich zu

$$\hat{\vartheta} = \left(\frac{(\mathbf{Z}^1)'\mathbf{Z}^1}{\hat{\sigma}_1^2} + \frac{(\mathbf{Z}^2)'\mathbf{Z}^2}{\hat{\sigma}_2^2} \right)^{-1} \left(\frac{(\mathbf{Z}^1)'\mathbf{y}^1}{\hat{\sigma}_1^2} + \frac{(\mathbf{Z}^2)'\mathbf{y}^2}{\hat{\sigma}_2^2} \right),$$

wobei $\mathbf{Z}^1$, $\mathbf{Z}^2$, $\mathbf{y}^1$ und $\mathbf{y}^2$ wie in Kapitel 4.2.1.2 definiert sind, also $\mathbf{Z}_t^1 = \mathbf{Z}_t$, falls $s_t = 1$, und 0 sonst, und analog für $\mathbf{Z}^2$, $\mathbf{y}^1$ und $\mathbf{y}^2$. (Dies ist äquivalent zu der Gleichung, die man erhält, wenn man in der Normalgleichung für ϑ die Parameter σ_1^2 und σ_2^2 durch $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$ ersetzt.)

Da $(\mathbf{Z}^i)'\mathbf{y}^i = (\mathbf{Z}^i)'\mathbf{Z}^i\hat{\vartheta}^i$, wobei $\hat{\vartheta}^i$ der OLS-Schätzer einer Regression für die Daten mit $s_t = i$ ist, gilt

$$\begin{aligned} \hat{\vartheta} &= \left(\frac{(\mathbf{Z}^1)'\mathbf{Z}^1}{\hat{\sigma}_1^2} + \frac{(\mathbf{Z}^2)'\mathbf{Z}^2}{\hat{\sigma}_2^2} \right)^{-1} \left(\frac{(\mathbf{Z}^1)'\mathbf{Z}^1\hat{\vartheta}^1}{\hat{\sigma}_1^2} + \frac{(\mathbf{Z}^2)'\mathbf{Z}^2\hat{\vartheta}^2}{\hat{\sigma}_2^2} \right) \\ &= (\mathbf{N}^1 + \mathbf{N}^2)^{-1}(\mathbf{N}^1\hat{\vartheta}^1 + \mathbf{N}^2\hat{\vartheta}^2), \end{aligned}$$

mit $$\mathbf{N}^i = \frac{(\mathbf{Z}^i)'\mathbf{Z}^i}{\hat{\sigma}_i^2} \text{ für } i = 1, 2.$$

Der Schätzer $\hat{\vartheta}$ ist also das mit den Matrizen $\mathbf{N}^i$ gewichtete Mittel der OLS-Schätzer aus den Regressionen über die Daten mit $s_t = i$. Die Matrix $\mathbf{N}^i$ ist die Inverse der geschätzten Varianz-Kovarianz-Matrix von $\hat{\vartheta}^i$ (ohne Bereinigung um die Freiheitsgrade), d.h. der Schätzer $\hat{\vartheta}^i$ mit der kleineren Varianz erhält ein größeres Gewicht in der Berechnung von $\hat{\vartheta}$.

In Anwendungen läßt sich $\hat{\vartheta}$ am leichtesten mit Hilfe einer Regression von

$$\tilde{\mathbf{y}} = \begin{pmatrix} \frac{y_1}{\hat{\sigma}_{s_1}} \\ \frac{y_2}{\hat{\sigma}_{s_2}} \\ \vdots \\ \frac{y_T}{\hat{\sigma}_{s_T}} \end{pmatrix} \quad \text{auf} \quad \tilde{\mathbf{Z}} = \begin{pmatrix} \frac{\mathbf{Z}_1}{\hat{\sigma}_{s_1}} \\ \frac{\mathbf{Z}_2}{\hat{\sigma}_{s_2}} \\ \vdots \\ \frac{\mathbf{Z}_T}{\hat{\sigma}_{s_T}} \end{pmatrix}$$

berechnen, d.h. $\hat{\vartheta}$ ist der WLS-(*Weighted Least Squares*-)Schätzer mit den Gewichten $\frac{1}{\hat{\sigma}_{s_t}}$, $t = 1,\ldots,T$.

Damit lauten die einzelnen Schritte des Oberhofer-Kmenta-Verfahrens:

0. Bestimme einen Startwert für $\hat{\vartheta}$, z.B. mit einer OLS-Schätzung.
1. Schätze damit die Fehlervarianzen durch

$$\hat{\sigma}_i^2 = \frac{1}{T_i}(\mathbf{y}^i - \mathbf{Z}^i\hat{\vartheta}^i)'(\mathbf{y}^i - \mathbf{Z}^i\hat{\vartheta}^i), i = 1, 2 .$$

2. Berechne damit den oben erläuterten FGLS-Schätzer $\hat{\vartheta}$.
3. Wiederhole Schritte 1. und 2., bis ein vorher gewähltes Konvergenzkriterium erfüllt ist.

In OBERHOFER und KMENTA [1974] und MAGNUS [1978] wird gezeigt, daß das Verfahren für diesen Fall an der Stelle des ML-Schätzers konvergiert.

Für den maximalen Wert der Log-Likelihood-Funktion gilt

$$\max(\log(f(\mathbf{y})) = -\frac{T}{2}(\log(2\pi)+1) - \frac{T_1}{2}\log\hat{\sigma}_1^2 - \frac{T_2}{2}\log\hat{\sigma}_2^2 \quad ,$$

wobei $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$ natürlich die ML-Schätzer darstellen. Das Verfahren läßt sich sehr leicht programmieren, wenn eine Routine für OLS-Schätzungen zur Verfügung steht. Es konvergierte in verschiedenen Testrechnungen[66] zuverlässig nach wenigen Iterationen, die Likelihood wuchs in allen Beispielen bei jedem Iterationsschritt.[67]

4.2.2 Unbekannter Schwellenwert

Bei einem Schwellenmodell mit unbekanntem Schwellenwert c muß dieser zusätzlich zu den übrigen Parametern aus den Daten geschätzt werden. Dabei tritt folgende Schwierigkeit auf: Die Likelihood weist Sprünge in c bei $z_0,...,z_T$ auf und ist konstant (in c) für Werte zwischen benachbarten Beobachtungen von z. Die Likelihood-Funktion ist also nicht nach c differenzierbar, und c ist lokal nicht identifiziert. In Anwendungen geht man deshalb wie folgt vor: Man berechnet die ML-Schätzer aus Kapitel 4.2.1 für die übrigen Parameter und den Wert der Likelihood für je einen festen Wert von c zwischen benachbarten Beobachtungen von z und gibt dann als Schätzung die Parameter, die der größten Likelihood insgesamt entsprechen, an.

Ähnlich kann man bei verallgemeinerten Schwellenmodellen mit unbekannten Schwellenwerten vorgehen. Hier führt man eine Gittersuche über beide unbekannten Schwellenwerte

[66] Alle Berechnungen in dieser Arbeit wurden mit dem Softwarepaket *RATS for Windows* (v. 4.30) der Firma *Estima* (Evanston, USA) durchgeführt.

[67] Für den Fall weiterer Regressoren $\mathbf{V}^1$ in Zustand 1 und $\mathbf{V}^2$ in Zustand 2 gilt wieder der Hinweis am Ende von Kapitel 4.2.1.2.

durch. Dabei tritt das Problem auf, daß die ersten Zustände nicht festgelegt sind, wenn die erste Beobachtung von z_t innerhalb des interessierenden Beobachtungszeitraums zwischen den beiden gewählten Schwellenparametern liegt.[68] In diesem Fall kann man die Gittersuche verfeinern, indem man die ML-Schätzer für beide möglichen Startwerte von s_t bestimmt.[69]

4.3 Mischungs- und Markov-Modelle

In der älteren Literatur herrschte die Ansicht vor, daß auch für Modelle, in denen der jeweils herrschende Zustand nicht beobachtet werden kann, die Aufteilung der Stichprobenwerte auf die verschiedenen Zustände aus den Daten bestimmt werden müsse.[70] Mittlerweile hat sich der vor allem in den Arbeiten von HAMILTON [1988 ff.] vertretene Standpunkt durchgesetzt, daß auch in diesem Fall nur die *Parameter* zu schätzen sind, und zwar nach Möglichkeit nach der ML-Methode. Rückschlüsse auf die Abfolge der Zustände lassen sich anhand der *geglätteten Wahrscheinlichkeiten* $P(s_t \mid \mathbf{y}, \mathbf{X}, \mathbf{W})$ ziehen, die aus den Parameterschätzern berechnet werden können.[71]

Im folgenden wird nach einigen einleitenden Bemerkungen über die stochastische Spezifikation des Zustandsprozesses die Likelihood-Funktion für Mischungs- und Markov-Modelle in verschiedenen Formen hergeleitet. Da Mischungsmodelle spezielle Markov-Modelle sind, wird zuerst auf letztere eingegangen und dann die Likelihood-Funktion für erstere als Spezialfall gewonnen.

Die maximale Laglänge sei wieder m. Die verzögert endogenen Variablen seien in den Regressoren $\mathbf{X}$ und $\mathbf{W}$ enthalten. Um die Ausführungen übersichtlich zu halten, werden folgende Notationen verwendet:

$$S_t = (s_t, s_{t-1}, \ldots, s_1),$$
$$Y_t = (y_t, y_{t-1}, \ldots, y_1, y_0, \ldots, y_{-m+1}).$$

Es ist also $S_T = \mathbf{s}$. Über den Zustandsprozeß werden folgende Annahmen gemacht: Die Folge s_t, $t = 1,\ldots,T$, ist eine (nicht notwendig homogene oder stationäre) Markov-Kette 1. Ordnung

[68] Dieses Problem kann auch bei verallgemeinerten Schwellenmodellen mit bekannten Schwellenwerten auftreten.

[69] Konsequenter wäre es, hier auch die Startverteilung des Zustandsprozesses durch einen weiteren freien Parameter zu beschreiben und diesen wie bei homogenen Markov-Modellen nach der ML-Methode aus den Daten zu schätzen. Damit verließe man aber den Rahmen, der für Schwellenmodelle üblich ist. Hier zeigt sich noch einmal die Verwandtschaft zwischen Markov- und verallgemeinerten Schwellenmodellen.

[70] Vgl. z.B. SCLOVE [1983] und TYSSEDAHL und TJØSTHEIM [1988].

[71] Vgl. dazu Kapitel 4.3.2.1, besonders S. 56, und Anhang A.

mit Zustandsraum {1, 2}. Die Übergangswahrscheinlichkeit $P(s_t|s_{t-1})$ wird vollständig von der exogenen Größe z_t determiniert:[72]

$$P(s_t|Y_{t-1},S_{t-1},\mathbf{X},\mathbf{W},\mathbf{z}) = P(s_t|s_{t-1},z_t).$$

Die Übergangswahrscheinlichkeit $P(s_t|s_{t-1},z_t)$ ist für jedes z_t eine (2×2)-Matrix mit zwei freien Größen. Wählt man dafür $p_t^{11}(z_t)$ und $p_t^{22}(z_t)$, so muß man für die Spezifikation des Modells noch geeignete funktionale Formen für diese Größen vorgeben, z.B. logistische Funktionen[73]

$$p_t^{11}(z_t) = \frac{e^{z_t\gamma_1+\delta_1}}{1+e^{z_t\gamma_1+\delta_1}} \quad \text{und} \quad p_t^{22}(z_t) = \frac{e^{z_t\gamma_2+\delta_2}}{1+e^{z_t\gamma_2+\delta_2}} .$$

Die freien Parameter sind in diesem Fall γ_1 , γ_2 , δ_1 , δ_2 und die Startwahrscheinlichkeit $P(s_1)$.[74] Ist s_t homogen, so gilt

$$p_t^{11}(z_t) \equiv p^{11} \text{ und } p_t^{22}(z_t) \equiv p^{22} \text{ für } t = 1,\ldots,T.$$

Die freien Parameter sind hier p^{11} und p^{22}. Ist die Kette nicht stationär, so kommt noch die Startwahrscheinlichkeit $P(s_1)$ hinzu.

Die Parameter dieser Modelle sind nicht strikt identifiziert: Vertauscht man alle Parameter eines Modells mit den jeweils entsprechenden des anderen Zustandes, so erhält man eines mit der gleichen Wahrscheinlichkeitsdichte, das sich für den Beobachter also nicht von dem ursprünglichen Modell unterscheiden läßt.[75] Diese Austauschbarkeit ist aber für empirische Anwendungen nicht bedeutsam: Sie läßt sich beseitigen durch eine geeignete Identifizierung eines der beiden Zustände etwa als denjenigen, für den ein bestimmter regimeabhängiger Parameter kleiner (oder größer) als der entsprechende des anderen Zustandes ist.[76]

Für eine beliebige Realisation des Zustandsvektors **s** eines inhomogenen *Markov-Modells*[77] gilt

---

[72] DIEBOLD, LEE und WEINBACH [1994] machen die Übergangswahrscheinlichkeit von z_{t-1} abhängig. Dies ist aber nur eine Frage der Definition von z_t.

[73] Unter den verschiedenen möglichen Funktionstypen hat die logistische Funktion bestimmte rechnerische Vorteile, die in Kapitel 4, Anhang B, erläutert werden, weswegen sie hier zugrunde gelegt wird. Eine Diskussion anderer Funktionstypen findet sich in LEE [1991].

[74] Statt der Größe **z** kann auch ein vektorieller Indikator $\mathbf{Z} = (\mathbf{z}_1,\ldots, \mathbf{z}_m)$ die Übergangswahrscheinlichkeiten bestimmen. Dann wären in allen Formeln $z_t\gamma_1$ und $z_t\gamma_2$ zu ersetzen durch $\mathbf{Z}_t\gamma_1$ und $\mathbf{Z}_t\gamma_2$ mit Vektoren $\gamma_1 = (\gamma_{1,1},\ldots, \gamma_{1,m})$ und $\gamma_2 = (\gamma_{2,1},\ldots, \gamma_{2,m})$.

[75] Die Likelihood-Funktion hat dementsprechend auch immer zwei Maxima gleicher Höhe an den korrespondierenden Stellen.

[76] Die Parameter sind also lokal identifiziert.

[77] Die entsprechenden Größen für homogene Modelle unterscheiden sich von den im folgenden angegebenen nur dadurch, daß der Zustandsindikator **z** nicht auftritt.

$$
\begin{aligned}
& f(\mathbf{y},\mathbf{s} \mid \mathbf{X},\mathbf{W},Y_0,\mathbf{z}) \\
&= f(Y_T,S_T \mid \mathbf{X},\mathbf{W},Y_0,\mathbf{z}) \\
&= f(y_T \mid Y_{T-1},S_T,\mathbf{X},\mathbf{W},\mathbf{z})P(s_T \mid Y_{T-1},S_{T-1},\mathbf{X},\mathbf{W},\mathbf{z}) \\
&\qquad \cdot f(Y_{T-1},S_{T-1} \mid \mathbf{X},\mathbf{W},Y_0,\mathbf{z}).
\end{aligned}
$$

Aufgrund der getroffenen stochastischen Annahmen lassen sich die ersten beiden Terme ersetzen, und man erhält

$$
\begin{aligned}
& f(\mathbf{y},\mathbf{s} \mid \mathbf{X},\mathbf{W},Y_0,\mathbf{z}) \\
&= f(y_T \mid s_T,\mathbf{X}_T,\mathbf{W}_T)P(s_T \mid s_{T-1},z_T)f(Y_{T-1},S_{T-1} \mid \mathbf{X},\mathbf{W},Y_0,\mathbf{z}).
\end{aligned}
$$

Löst man den letzten Term sukzessive auf diese Weise auf, so erhält man schließlich

$$
\begin{aligned}
& f(\mathbf{y},\mathbf{s} \mid \mathbf{X},\mathbf{W},Y_0,\mathbf{z}) \\
&= \left(\prod_{t=2}^{T} f(y_t \mid s_t,\mathbf{X}_t,\mathbf{W}_t)P(s_t \mid s_{t-1},z_t)\right) f(y_1 \mid s_1,\mathbf{X}_1,\mathbf{W}_1)P(s_1 \mid Y_0,\mathbf{X},\mathbf{W},\mathbf{z}).
\end{aligned}
$$

Alle Terme bis auf den letzten lassen sich durch die Parameter und die beobachteten Größen ausdrücken. Anstatt den letzten Term noch durch (Summen von) Übergangswahrscheinlichkeiten, bedingte Dichten von $\mathbf{y}$ und die Startwahrscheinlichkeiten $P(s_1)$ auszudrücken, ist es auch möglich, die *bedingte Startwahrscheinlichkeit* $P(s_1 = 1|Y_0,\mathbf{X},\mathbf{W},\mathbf{z})$ direkt als freien Parameter zu verwenden. Sei also

$$
\begin{pmatrix} \rho^1 \\ 1-\rho^1 \end{pmatrix} = \begin{pmatrix} P(s_1 = 1|Y_0,\mathbf{X},\mathbf{W},\mathbf{z}) \\ P(s_1 = 2|Y_0,\mathbf{X},\mathbf{W},\mathbf{z}) \end{pmatrix}.
$$

Damit läßt sich der obige Ausdruck für $f(\mathbf{y},\mathbf{s}|\mathbf{X},\mathbf{W},Y_0,\mathbf{z})$ vollständig aus den Daten und den Parametern berechnen. Die gesuchte Likelihood-Funktion ergibt sich jetzt, indem man die Summe aller dieser Ausdrücke für verschiedene Realisationen des Zustandsvektors $\mathbf{s}$ bildet:

$$
\begin{aligned}
& f(\mathbf{y}|\mathbf{X},\mathbf{W},Y_0,\mathbf{z}) \\
&= \sum_{\mathbf{s}} f(\mathbf{y},\mathbf{s}|\mathbf{X},\mathbf{W},Y_0,\mathbf{z}) \\
&= \sum_{\mathbf{s}} \left[\left(\prod_{t=2}^{T} f(y_t|s_t,\mathbf{X}_t,\mathbf{W}_t)P(s_t|s_{t-1},z_t)\right) f(y_1|s_1,\mathbf{X}_1,\mathbf{W}_1)P(s_1|Y_0,\mathbf{X},\mathbf{W},\mathbf{z}) \right].
\end{aligned}
$$

Dieser geschlossene Ausdruck für die Likelihood-Funktion ist für theoretische Überlegungen, nicht aber für die praktische Ermittlung des Wertes der (Log-)Likelihood-Funktion an speziellen Parameterwerten geeignet, da zu seiner Berechnung die Summation über alle möglichen 2^T Zustände erforderlich wäre, so daß auch für moderate Zeitreihenlängen die Anzahl der nötigen Rechenschritte selbst mit schnellen Computern sehr hohe Rechenzeiten ergäbe. Es

gibt aber Verfahren, den Wert der (Log-)Likelihood-Funktion mit deutlich weniger Rechenaufwand zu ermitteln, z.B. das rekursive Verfahren von COSSLETT und LEE [1985]:[78] Sei

$$Q_t(s_t) = f(Y_t, s_t) \ .$$

Es gilt

$$\begin{aligned} Q_t(s_t) &= \sum_{s_{t-1}} f(Y_t, s_t, s_{t-1}) \\ &= \sum_{s_{t-1}} f(y_t, s_t \mid Y_{t-1}, s_{t-1}) Q_{t-1}(s_{t-1}) \\ &= \sum_{s_{t-1}} f(y_t \mid Y_{t-1}, s_t, s_{t-1}) P(s_t \mid Y_{t-1}, s_{t-1}) Q_{t-1}(s_{t-1}) \\ &= \sum_{s_{t-1}} f(y_t \mid Y_{t-1}, s_t) P(s_t \mid s_{t-1}) Q_{t-1}(s_{t-1}) \ . \end{aligned}$$

Dieser Ausdruck enthält neben leicht aus den Daten berechenbaren Größen nur den Wert $Q_{t-1}(s_{t-1})$, der im letzten Rekursionsschritt berechnet wurde. Die Rekursion beginnt mit

$$Q_1(s_1) = f(y_1, s_1) = f(y_1 \mid s_1) P(s_1)$$

und endet mit der Berechnung von $Q_T(s_T)$, woraus man schließlich das gesuchte

$$f(\mathbf{y}) = f(y_T) = \sum_{s_T} f(y_T, s_T) = \sum_{s_T} Q_T(s_T)$$

erhält.

Eine Alternative zu diesem Verfahren ergibt sich aus der für jede Zeitreihe gültigen Beziehung

$$f(\mathbf{y}) = \left(\prod_{t=2}^{T} f(y_t | Y_{t-1}) \right) f(y_1) \ .$$

Es gilt

$$f(y_1) = \sum_{s_1} f(y_1 | s_1) \ ,$$

und die Größen $f(y_t \mid Y_{t-1})$ lassen sich zum Beispiel mit dem in Anhang A beschriebenen *Filter* berechnen, der Teil des in Kapitel 4.3.2 erläuterten EM-Algorithmus ist.[79]

Inhomogene *Mischungsmodelle* ergeben sich aus inhomogenen Markov-Modellen durch die Restriktionen $\gamma_1 = -\gamma_2$ und $\delta_1 = -\delta_2$, falls die Übergangswahrscheinlichkeiten logistische Funktionen von z_t sind. Es gilt dann nämlich

[78] Für den Rest des Kapitels werden die vorherbestimmten Größen in der Notation nicht mehr berücksichtigt.

[79] Vgl. hierzu auch GRAY [1996], Appendix.

$$p_t^{11}(z_t) = \frac{e^{\gamma_1 z_t + \delta_1}}{1 + e^{\gamma_1 z_t + \delta_1}} = \frac{e^{-\gamma_2 z_t + \delta_2}}{1 + e^{-\gamma_2 z_t + \delta_2}} = \frac{1}{1 + e^{\gamma_1 z_t + \delta_1}}$$
$$= 1 - \frac{e^{\gamma_2 z_t + \delta_2}}{1 + e^{\gamma_2 z_t + \delta_2}} = 1 - p_t^{22}(z_t) = p_t^{12}(z_t).$$

Sinnvollerweise wird man in diesem Fall $P(s_1 = 1|Y_0, \mathbf{X}, \mathbf{W}, \mathbf{z})$ nicht mehr als freien Parameter modellieren, sondern die zusätzliche Restriktion

$$P(s_1 = 1|Y_0, \mathbf{X}, \mathbf{W}, \mathbf{z}) = p_t^{11}(z_1)$$

einführen, so daß für alle $t = 1,...,T$ gilt

$$P(s_t = 1|z_t) = p_t^{11}(z_t) \ .$$

Das Modell verwendet dann die zwei freien Parameter γ_1 und δ_1 für den Zustandsprozeß. Es gilt hier

$$f(\mathbf{y}|\mathbf{X}, \mathbf{W}, Y_0, \mathbf{z})$$
$$= \sum_{\mathbf{s}} \left(\prod_{t=1}^{T} f(y_t|s_t, \mathbf{X}_t, \mathbf{W}_t) P(s_t|z_t) \right)$$
$$= \prod_{t=1}^{T} \left(f(y_t|s_t = 1, \mathbf{X}_t, \mathbf{W}_t) P(s_t = 1|z_t) + f(y_t|s_t = 2, \mathbf{X}_t, \mathbf{W}_t) P(s_t = 2|z_t) \right)$$

Eine derartige Produktdarstellung (bzw. Summendarstellung für die Log-Likelihood), die die Arbeit mit Mischungsmodellen sehr vereinfacht, existiert für (echte) Markov-Modelle nicht.[80]

Da die Normalgleichungen für Mischungs- und Markov-Modelle nicht linear sind, läßt sich kein geschlossener Ausdruck für die ML-Schätzer als Funktion der beobachteten Größen angeben. Deswegen muß man auf iterative Verfahren zurückgreifen, die, beginnend mit vorgegebenen Startwerten (die eventuell mit Hilfe eines Gittersuchverfahrens gewonnen werden), nach einem vorgegebenen Algorithmus schrittweise neue Parameterwerte berechnen, bis ein vordefiniertes Abbruchkriterium erfüllt ist. Im folgenden wird eine kurze Übersicht über herkömmliche Iterationsverfahren, die hier als *numerische Verfahren* bezeichnet werden, gegeben. Anschließend wird eine weitere Methode – der sogenannte *EM-Algorithmus* -, die sich in Anwendungen als besonders vorteilhaft erwiesen hat, erläutert.

4.3.1 Numerische Maximierung der Likelihood

Die meisten herkömmlichen Verfahren[81] benutzen (analytisch oder numerisch berechnete) Ableitungen 1. und 2. Ordnung für die Berechnung der jeweils folgenden Parameterwerte.

80 GOLDFELD und QUANDT [1973] gingen noch von einer solchen Produktdarstellung aus. Ihr Ausdruck für die Log-Likelihood-Funktion (Gleichung (13), S. 8) ist aber falsch.

81 Ein Überblick über numerische Verfahren findet sich z.B. in JUDGE u.a. [1985], Appendix B, oder HAMILTON [1994b], Kap. 5.7.

Man bezeichnet sie deshalb als *Gradientenverfahren*. Sie benutzen die Ableitungen der Log-Likelihood-Funktion, um eine Richtung mit wachsendem Funktionswert zu bestimmen, in der der nächste Parameterwert gewählt wird. Am häufigsten verwendet werden das Newton-Raphson-, das Gauss-, das DFP-(Davidson-Fletcher-Powell-) und das BHHH-(Bernd-Hall-Hall-Hausmann-)Verfahren und verschiedene Varianten davon.

Ohne Ableitungen kommt der *Simplex-Algorithmus* aus, bei dem zur Bestimmung der nächsten Parameterwerte die Werte der Likelihood-Funktion an den Ecken eines Simplexes im Parameterraum um den letzten Parameterwert bestimmt werden. Bei diesem Algorithmus wird in jedem Schritt die Ecke mit dem niedrigsten Funktionswert durch die Ecke ersetzt, die man erhält, wenn man sie an der Hyperebene durch die übrigen Ecken spiegelt.

Ein Vorteil der Gradientenverfahren, für die es in vielen Softwarepaketen vorgefertigte Prozeduren gibt, ist, daß sie an einem Konvergenzpunkt, an dem die Iteration abgebrochen wird, automatisch Schätzer für die asymptotische Varianz-Kovarianz-Matrix der Parameter liefern, vgl. z.B. HAMILTON [1994b], S. 143 f., oder GOURIEROUX und MONFORT [1995a], S. 185 f. Diese können für asymptotische Tests von Hypothesen über die Parameter verwendet werden.

Für Mischungsverteilungen lassen sich solche Verfahren in den gängigen Softwarepaketen leicht implementieren, da die Likelihood-Funktion für sie eine einfache Produktgestalt hat. Für Markovmodelle, bei denen solch eine Darstellung nicht existiert, kann man auf eine rekursive Darstellung der Likelihood-Funktion zurückgreifen, wenn die verwendete Maximierungsprozedur rekursive Programmierung gestattet.[82]

4.3.2 Der EM-Algorithmus

Eine Alternative zu numerischen Verfahren ist der EM-(*Expectation-Maximisation*-) Algorithmus. Er stellt ein iteratives Verfahren dar, mit dem ML-Schätzer für Modelle berechnet werden können, in denen die Wahrscheinlichkeitsdichte beobachteter Variablen von unbeobachteten (*latenten*) Größen (wie hier dem Zustandsprozeß **z**) abhängt, so daß die vollständige Likelihood-Funktion ein Integral (oder eine Summe) solcher Wahrscheinlichkeitsdichten über alle möglichen Realisierungen der unbeobachteten Größen ist.[83] Der EM-Algorithmus in allgemeiner Form wurde, aufbauend u.a. auf BAUM u.a. [1970], von

[82] Wie etwa die Prozeduren *maximize* in RATS oder *maxlik* in GAUSS.

[83] Gelegentlich wird der EM-Algorithmus als Alternative zu ML-Verfahren bezeichnet. Dies ist nicht korrekt: Der EM-Algorithmus ist eine Methode zur Berechnung der ML-Schätzer, ebenso wie die hier als *numerische Verfahren* bezeichneten, mehrheitlich auf der Verwendung von Ableitungen der (Log-)Likelihood-Funktion basierenden Algorithmen.

DEMPSTER, LAIRD und RUBIN [1977] eingeführt.[84] Sie identifizierten verschiedene zu diesem Zeitpunkt bekannte Verfahren zur Bestimmung von ML-Schätzern in Modellen mit latenten Variablen als Spezialfälle des EM-Algorithmus. Außer für Mischungs- und Markov-Modelle wird der EM-Algorithmus in der ökonometrischen Praxis noch für die Schätzung anderer Modelltypen verwendet, darunter insbesondere für Zustandsraummodelle.[85] Hier stellt der Zustandsvektor die latente Größe dar.

Die Größen **Y** seien beobachtbar, nicht jedoch die latenten Variablen **S**. Das Ziel des Verfahrens ist, die Likelihood-Funktion

$$\log f(\mathbf{Y};\lambda) = \log \int_S f(\mathbf{Y},\mathbf{S};\lambda)d\mathbf{S}$$

zu maximieren, wobei λ die unbekannten Parameter des Modells darstellt.[86] Anstatt dieses Integral zu maximieren, werden Integration und Maximierung im EM-Algorithmus gewissermaßen in zwei Schritte aufgespalten, den E-Schritt und den M-Schritt:

Im *E-Schritt* wird die momentan "beste" Schätzung des Parameters, $\lambda^{(n)}$, verwendet, um den bedingten Erwartungswert von $\log f(\mathbf{Y},\mathbf{S};\lambda)$ (der *Complete-Data Log-Likelihood*) gegeben **Y** zu berechnen. Sei also

$$Q(\lambda \mid \mathbf{Y};\lambda^{(n)}) = E(\log f(\mathbf{Y},\mathbf{S};\lambda) \mid \mathbf{Y};\lambda^{(n)}) = \int_S (\log f(\mathbf{Y},\mathbf{S};\lambda)) f(\mathbf{S} \mid \mathbf{Y};\lambda^{(n)})d\mathbf{S} \ .$$

Im *M-Schritt* wird der neue Schätzer für λ als der Wert berechnet, der diesen Erwartungswert maximiert, d.h.

$$\lambda^{(n+1)} = \arg\max_\lambda Q(\lambda \mid \mathbf{Y};\lambda^{(n)}) \ .$$

Statt $\log f(\mathbf{Y};\lambda)$ wird also gewissermaßen die zum momentanen Kenntnisstand ($\mathbf{Y}$, $\lambda^{(n)}$) plausibelste Approximation (die bedingte Erwartung) maximiert.

HAMILTON [1991] maximiert statt dessen

$$Q'(\lambda \mid \mathbf{Y};\lambda^{(n)}) = \int_S (\log f(\mathbf{Y},\mathbf{S};\lambda)) f(\mathbf{S},\mathbf{Y};\lambda^{(n)})d\mathbf{S} \ .$$

Dies führt auf den gleichen Algorithmus: Es gilt

$$Q'(\lambda \mid \mathbf{Y};\lambda^{(n)}) = f(\mathbf{Y};\lambda^{(n)})Q(\lambda \mid \mathbf{Y};\lambda^{(n)}) \ ,$$

deswegen nehmen beide Funktionen ihr Maximum bei demselben Wert für λ an.

---

[84] Vgl. auch RUUD [1991] und GOURIEROUX und MONFORT [1995a], Kap. 13.3.

[85] Vgl. hierzu ENGLE und WATSON [1983].

[86] Ist **S** diskret, so bedeutet das Integral in obigem Ausdruck die Summe über alle möglichen Werte von **S**.

Der EM-Algorithmus besteht nun darin, daß erst ein Startwert $\lambda^{(0)}$ gewählt und damit durch abwechselnde Ausführung von E- und M-Schritten eine Folge von Parametern $\lambda^{(0)}, \lambda^{(1)}, \lambda^{(2)}...$ ermittelt wird. Diese Sequenz hat die folgende wichtige Eigenschaft:[87]

$$\log f(\mathbf{Y};\lambda^{(n+1)}) \geq \log f(\mathbf{Y};\lambda^{(n)}) \ .$$

Solange, bis die Parameterfolge konvergiert, führt also jeder Iterationsschritt zu einem höheren (oder wenigstens zu dem gleichen) Wert der (Log-)Likelihood. Weiter gilt, daß, falls die Folge konvergiert, der Grenzwert die Normalgleichungen der (Log-)Likelihood-Funktion erfüllt.[88]

In Anwendungen kann man also lokale Maxima der (Log-)Likelihood-Funktion näherungsweise bestimmen, indem man geeignete Startwerte bestimmt und dann mit dem EM-Algorithmus solange iterativ weitere Parameterwerte berechnet, bis ein vordefiniertes Konvergenzkriterium erfüllt ist, z.B. bis die letzte absolute Veränderung der Parameter (in der euklidischen oder einer anderen Norm gemessen) oder der (Log-)Likelihood-Funktion eine vorgegebene Schwelle unterschreitet.

Das Verfahren ist dann vorteilhaft, wenn sich $Q(\lambda \mid \mathbf{Y};\lambda^{(n)})$ im Vergleich zur Log-Likelihood-Funktion leicht maximieren läßt, wie es häufig (und besonders für Markov-Modelle) der Fall ist.

Daß man bei diesem Verfahren das *globale* Maximum findet, ist allerdings nicht sichergestellt. Dieses Problem teilt der EM-Algorithmus mit anderen iterativen Verfahren zur Bestimmung des Maximums der Log-Likelihood-Funktion. Empirische Anwendungen haben aber übereinstimmend ergeben, daß der EM-Algorithmus sehr *stabil* ist, d.h. daß er unabhängig von der Wahl der Startwerte in den allermeisten Fällen den gleichen Konvergenzpunkt erreicht, während andere Iterationsverfahren (je nach Startwert) häufig an Punkten mit niedrigerem Wert der Log-Likelihood konvergieren.[89]

Im folgenden wird die spezielle Form des EM-Algorithmus für Markov- und Mischungsmodelle erläutert. Der EM-Algorithmus für Markov-Modelle mit exogenen Regressoren wurde zuerst von LINDGREN [1978] präsentiert.[90] Diese Arbeit fand jedoch in der Folgezeit kaum

[87] Vgl. DEMPSTER, LAIRD und RUBIN [1977], S. 7, *Theorem* 1, und HAMILTON [1990], S. 48, *Observation* 1.

[88] Vgl. HAMILTON [1990], S. 49, *Observation* 2.

[89] Der EM-Algorithmus ist in der Nähe eines Konvergenzpunktes verhältnismäßig langsam (d.h. er benötigt dort viele Iterationen für geringe Veränderungen der Parameterwerte), weshalb manche Autoren empfehlen, nur die ersten Iterationsschritte mit dem EM-Algorithmus zu berechnen und den Durchlauf dann mit einem anderen Iterationsverfahren (schneller) zu beenden.

[90] KIEFER [1980] stellte, aufbauend auf BEHBOODIAN [1970], aber offensichtlich ohne Kenntnis der Arbeiten von DEMPSTER, LAIRD und RUBIN [1977] und LINDGREN [1978], einen iterativen Algorithmus zur Bestimmung der

Beachtung, bis HAMILTON [1990] ihn auf den Fall von Zeitreihenmodellen mit zugrunde liegender homogener Markov-Kette verallgemeinerte. Die erste Darstellung für inhomogene Markov-Modelle (ohne autoregressive Dynamik oder exogene Größen) stammt von DIEBOLD, LEE und WEINBACH [1994].

4.3.2.1 Der E-Schritt

Die im folgenden angegebenen Wahrscheinlichkeitsdichten sind wieder als "gegeben **X**, **W** und **z**" zu verstehen, was nicht mehr ausdrücklich notiert wird. Dagegen ist es hier nötig, den Parameter anzugeben, an dem die Likelihood-Terme ausgewertet werden. Dafür werden alle Parameter des jeweiligen Modells (Regressionsparameter und solche, die die Verteilung des Zustandsprozesses festlegen), in dem Vektor λ zusammengefaßt.

Für Markov-Modelle gilt

$$\log f(\mathbf{y},\mathbf{s};\lambda) = \sum_{t=1}^{T} \log f(y_t \mid s_t;\lambda) + \sum_{t=2}^{T} \log P(s_t \mid s_{t-1};\lambda) + \log P(s_1;\lambda).$$

Daraus erhält man

$$\begin{aligned}
&Q(\lambda \mid \mathbf{y};\lambda^{(n)}) \\
&= E\left(\log f(\mathbf{y},\mathbf{s};\lambda) \mid \mathbf{y};\lambda^{(n)}\right) \\
&= \sum_{\mathbf{s}} \left(\sum_{t=1}^{T} \log f(y_t \mid s_t;\lambda) + \sum_{t=2}^{T} \log P(s_t \mid s_{t-1};\lambda) + \log P(s_1;\lambda) \right) P(\mathbf{s} \mid \mathbf{y};\lambda^{(n)}) \\
&= \sum_{t=1}^{T} \left(\sum_{\mathbf{s}} \log f(y_t \mid s_t;\lambda) P(\mathbf{s} \mid \mathbf{y};\lambda^{(n)}) \right) \\
&+ \sum_{t=2}^{T} \left(\sum_{\mathbf{s}} \log P(s_t \mid s_{t-1};\lambda) P(\mathbf{s} \mid \mathbf{y};\lambda^{(n)}) \right) + \sum_{\mathbf{s}} \log P(s_1;\lambda) P(\mathbf{s} \mid \mathbf{y};\lambda^{(n)}).
\end{aligned}$$

In jedem Summanden dieser Summe hängt der erste Faktor nur von einem bzw. zwei Zuständen s_t ab. Deshalb kann man die Summation über die jeweils anderen Zustände sofort ausführen und erhält

$$\begin{aligned}
Q(\lambda \mid \mathbf{y};\lambda^{(n)}) &= \sum_{t=1}^{T} \left(\sum_{s_t} \log f(y_t \mid s_t;\lambda) P(s_t \mid \mathbf{y};\lambda^{(n)}) \right) \\
&+ \sum_{t=2}^{T} \left(\sum_{s_t, s_{t-1}} \log P(s_t \mid s_{t-1};\lambda) P(s_t, s_{t-1} \mid \mathbf{y};\lambda^{(n)}) \right) + \sum_{s_1} \log P(s_1;\lambda) P(s_1 \mid \mathbf{y};\lambda^{(n)}).
\end{aligned}$$

Für Mischungsmodelle ergibt sich analog

ML-Schätzer für homogene Mischungsmodelle vor. Dies *ist* der EM-Algorithmus für diese Modelle in einer etwas ungewöhnlichen Formulierung.

$$Q(\lambda \mid \mathbf{y};\lambda^{(n)}) = \sum_{t=1}^{T}\left(\sum_{s_t}\left(\log f(y_t \mid s_t;\lambda) + \log P(s_t;\lambda)\right)P(s_t \mid \mathbf{y};\lambda^{(n)})\right).$$

Die im M-Schritt zu maximierende Funktion $Q(\lambda \mid \mathbf{y};\lambda^{(n)})$ besteht demnach aus den Termen, die die Log-Likelihood-Funktion bilden, gewichtet mit den bedingten Wahrscheinlichkeiten, mit denen die Zustandsvariable(n) s_t die jeweiligen Werte angenommen hat. Für Markov-Modelle ist $Q(\lambda \mid \mathbf{y};\lambda^{(n)})$ die Summe von $2T + 4(T-1) + 2 = 6T - 2$ Produkten aus je 2 Faktoren, was eine deutliche Vereinfachung gegenüber der Likelihood-Funktion darstellt, die aus einer Summe von 2^T Produkten aus je $2T$ Faktoren besteht. Für Mischungsmodelle sind $Q(\lambda \mid \mathbf{y};\lambda^{(n)})$ und die Produktform der Likelihood etwa gleich komplex.

Für den E-Schritt werden demzufolge die bedingten oder *geglätteten* Wahrscheinlichkeiten[91] der Zustandsvariablen

$$P(s_t \mid \mathbf{y};\lambda^{(n)})$$

und für Markov-Modelle auch die bedingten gemeinsamen Wahrscheinlichkeiten aufeinanderfolgender Zustandsvariablen

$$P(s_t, s_{t-1} \mid \mathbf{y};\lambda^{(n)})$$

benötigt. Insbesondere die geglätteten Wahrscheinlichkeiten sind unabhängig vom EM-Algorithmus interessant: Ist $\lambda^{(n)}$ ein Schätzer für λ (etwa der ML-Schätzer), gestatten sie Rückschlüsse von den vorliegenden Informationen (den beobachteten Größen) auf die unbeobachtbaren Zustände. Ist z.B. $P(s_t = 1 \mid \mathbf{y};\lambda^{(n)})$ groß (in der Nähe von 1), so wird man mit großer Wahrscheinlichkeit davon ausgehen können, daß $s_t = 1$ gegolten hat und umgekehrt. Die geglätteten Wahrscheinlichkeiten werden aber auch zur Bestimmung von Prognosen mit minimalem mittleren quadratischen Prognosefehler benötigt.[92]

Für Mischungsmodelle ohne verzögert endogene Regressoren lassen sich diese Größen leicht berechnen. Hier gilt

[91] HAMILTON [1990] bezeichnet sie als *smoothed probabilities*. Dieser Ausdruck hat sich seitdem in der Literatur durchgesetzt.

[92] Vgl. hierzu Kapitel 8.2.

$$
\begin{aligned}
P(s_t \mid Y_T) &= \frac{f(s_t, Y_T)}{f(Y_T)} = \frac{\sum_{s_1,\ldots,s_{t-1},s_{t+1},\ldots,s_T} f(Y_T, s_T)}{f(Y_T)} \\
&= \sum_{s_1,\ldots,s_{t-1},s_{t+1},\ldots,s_T} \frac{\prod_{t'=1}^{T} f(y_{t'} \mid s_{t'}, Y_{t'-1}) P(s_{t'})}{\prod_{t'=1}^{T} f(y_{t'} \mid Y_{t'-1})} \\
&= \frac{\prod_{\substack{t'=1,\ldots,T \\ t' \neq t}} \left(f(y_{t'}, s_{t'} = 1 \mid Y_{t'-1}) + f(y_{t'}, s_{t'} = 2 \mid Y_{t'-1}) \right) f(y_t \mid s_t, Y_{t-1}) P(s_t)}{\prod_{t'=1}^{T} f(y_{t'} \mid Y_{t'-1})} \\
&= \frac{\prod_{\substack{t'=1,\ldots,T \\ t' \neq t}} \left(f(y_{t'} \mid Y_{t'-1}) \right) f(y_t \mid s_t, Y_{t-1}) P(s_t)}{\prod_{t'=1}^{T} f(y_{t'} \mid Y_{t'-1})} \\
&= \frac{f(y_t \mid s_t, Y_{t-1}) P(s_t)}{f(y_t \mid Y_{t-1})} = \frac{f(y_t \mid s_t, Y_{t-1}) P(s_t)}{\sum_{s_{t'}=1,2} f(y_t \mid s_{t'}, Y_{t-1}) P(s_{t'})}.
\end{aligned}
$$

In dem letzten Ausdruck stehen nur Größen, die sich direkt aus den Daten berechnen lassen. Diese Ableitung macht entscheidend von der Produktdarstellung der Likelihood Gebrauch. Da eine vergleichbare Darstellung für Markov-Modelle nicht existiert, ist die Berechnung der geglätteten Zustandswahrscheinlichkeiten in diesem Fall ungleich komplizierter als für Mischungsmodelle. HAMILTON [1990] hat ein rekursives Verfahren angegeben, mit dem die gesuchten Größen für Markov-Zeitreihenmodelle berechnet werden können.[93] Es wird im Anhang A zu diesem Kapitel in einer leicht modifizierten Form wiedergegeben, die für Modelle ohne autoregressive Dynamik geeigneter ist. Die Darstellung lehnt sich an DIEBOLD, LEE und WEINBACH [1994] an und korrigiert einen Fehler in ihrem Algorithmus.

Hamilton nutzte in seiner Herleitung des E-Schritts des EM-Algorithmus die Verwandtschaft zwischen Markov- und Zustandsraummodellen (vgl. Kapitel 2.4.2). Deswegen ist es nicht verwunderlich, daß eine starke Analogie zwischen dem *Kalman-Filter*, mit dem auf den Wert des Zustandsvektors eines Zustandsraummodells für gegebene Werte der beobachtbaren Variablen und der Parameter geschlossen werden kann, und dem E-Schritt des EM-Algorithmus besteht.[94]

[93] Vgl. hierzu auch KIM [1994].

[94] Vgl. hierzu ENGLE und WATSON [1983].

4.3.2.2 Der M-Schritt

Im M-Schritt ist der Wert für λ gesucht, der

$$Q(\lambda \mid \mathbf{y};\lambda^{(n)}) = \sum_{t=1}^{T}\left(\sum_{s_t}\left(\log f(y_t \mid s_t, \mathbf{X}_t, \mathbf{W}_t;\lambda) + \log P(s_t \mid z_t;\lambda)\right)P(s_t \mid \mathbf{y};\lambda^{(n)})\right)$$

(für Mischungsmodelle) bzw.

$$\begin{aligned} Q(\lambda \mid \mathbf{y};\lambda^{(n)}) = &\sum_{t=1}^{T}\left(\sum_{s_t}\log f(y_t \mid s_t, \mathbf{X}_t, \mathbf{W}_t;\lambda)P(s_t \mid \mathbf{y};\lambda^{(n)})\right) \\ &+\sum_{t=2}^{T}\left(\sum_{s_t,s_{t-1}}\log P(s_t \mid s_{t-1}, z_t;\lambda)P(s_t, s_{t-1} \mid \mathbf{y};\lambda^{(n)})\right) \\ &+\sum_{s_1}\log P(s_1;\lambda)P(s_1 \mid \mathbf{y};\lambda^{(n)}) \end{aligned}$$

(für Markov-Modelle) maximiert. Dieser wird bestimmt, indem die zugehörigen Normalgleichungen nach diesen Parametern aufgelöst werden.

4.3.2.2.1 Die Normalgleichungen für die Regressionsparameter und die Störvarianz(en)

Die Normalgleichungen für α, β_1, β_2, σ_1 und σ_2 bzw. σ erhält man durch Differentiation von $Q(\lambda \mid \mathbf{y};\lambda^{(n)})$ nach diesen Parametern, die nur in den Termen

$$\log f(y_t \mid s_t, \mathbf{X}_t, \mathbf{W}_t;\lambda) = -\frac{1}{2}\log(2\pi) - \log(\sigma_{s_t}) - \frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_{s_t})^2}{2\sigma_{s_t}^2}$$

(für $t = 1,\ldots, T$) auftreten. Man erhält z.B. folgende Normalgleichung für β_1:

$$\frac{\partial Q(\lambda \mid \mathbf{y};\lambda^{(n)})}{\partial \beta_1} = \mathbf{0} \quad \Leftrightarrow$$

$$\sum_{t=1}^{T}\mathbf{X}_t'(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)P(s_t = 1 \mid \mathbf{y},\lambda^{(n)}) = \mathbf{0}\ . \qquad \text{(a1)}$$

Diese und die übrigen Gleichungen sind völlig analog zu den Normalgleichungen für Schwellenmodelle aus Kapitel 4.2.1. Die dort auftretenden Terme für die beiden Zustände werden hier gewichtet mit den geglätteten Wahrscheinlichkeiten für diese Zustände. Bis auf diesen Unterschied entspricht der M-Schritt des EM-Algorithmus im wesentlichen der Berechnung von ML-Schätzern für beobachtbare Zustände. Anders gesagt: Ersetzt man in den Normalgleichungen des M-Schrittes die geglätteten Wahrscheinlichkeiten $P(s_t = 1 \mid \mathbf{y};\lambda^{(n)})$ durch

$$D_c(z_t) = \begin{cases} 1 \text{ falls } z_t < c \\ 0 \text{ falls } z_t \geq c \end{cases}$$

und $P(s_t = 2 \mid \mathbf{y}, \lambda^{(n)})$ durch $1 - D_c(z_t)$, so erhält man die Gleichungen aus Kapitel 4.2.1. Dieser Zusammenhang resultiert aus der Tatsache, daß Schwellenmodelle spezielle inhomogene Markov- bzw.- Mischungsmodelle sind (nämlich solche, bei denen die Übergangs- bzw. die Aufenthaltswahrscheinlichkeiten $p_t^{11}(z_t)$ und $p_t^{22}(z_t)$ Stufenfunktionen von z_t sind).[95]

4.3.2.2.2 Berechnung der neuen Regressionsparameter und Störvarianz(en) durch Hilfsregressionen

Der M-Schritt läßt sich ähnlich wie eine ML-Schätzung für Schwellenmodelle mit Hilfe von Einfachregressionen berechnen. Da hier aber Zeitreihen de Länge $2T$ verwendet werden müssen, unterscheiden sich die Formeln von denen aus Kapitel 4.2.1.2.

Sei $\mathbf{X}^i$ die Matrix, deren t-te Zeile $\mathbf{X}_t^i$ gleich $\mathbf{X}_t \cdot \sqrt{P(s_t = i \mid \mathbf{y}, \lambda^{(n)})}$ $(i = 1, 2)$. Analog seien $\mathbf{W}^1$, $\mathbf{W}^2$, $\mathbf{y}^1$ und $\mathbf{y}^2$ definiert. Mit diesen Bezeichnungen lauten die Normalgleichungen

$$\sum_{t=1}^{T} (\mathbf{X}_t^1)'(y_t^1 - \mathbf{W}_t^1\alpha - \mathbf{X}_t^1\beta_1) = \mathbf{0} \ , \tag{a1}$$

$$\sum_{t=1}^{T} (\mathbf{X}_t^2)'(y_t^2 - \mathbf{W}_t^2\alpha - \mathbf{X}_t^2\beta_2) = \mathbf{0} \ , \tag{a2}$$

$$\sum_{t=1}^{T} \frac{(\mathbf{W}_t^1)'(y_t^1 - \mathbf{W}_t^1\alpha - \mathbf{X}_t^1\beta_1)}{\sigma_1^2} + \sum_{t=1}^{T} \frac{(\mathbf{W}_t^2)'(y_t^2 - \mathbf{W}_t^2\alpha - \mathbf{X}_t^2\beta_2)}{\sigma_2^2} = 0 \ , \tag{b}$$

$$\sum_{t=1}^{T} (\mathbf{W}_t^1)'(y_t^1 - \mathbf{W}_t^1\alpha - \mathbf{X}_t^1\beta_1) + \sum_{t=1}^{T} (\mathbf{W}_t^2)'(y_t^2 - \mathbf{W}_t^2\alpha - \mathbf{X}_t^2\beta_2) = 0 \ , \tag{b'}$$

$$\sigma_1^2 = \frac{\sum_{t=1}^{T} (y_t^1 - \mathbf{W}_t^1\alpha - \mathbf{X}_t^1\beta_1)^2}{\sum_{t=1}^{T} P(s_t = 1 \mid \mathbf{y}, \lambda^{(n)})} \ , \tag{c1}$$

$$\sigma_2^2 = \frac{\sum_{t=1}^{T} (y_t^2 - \mathbf{W}_t^2\alpha - \mathbf{X}_t^2\beta_2)^2}{\sum_{t=1}^{T} P(s_t = 2 \mid \mathbf{y}, \lambda^{(n)})} \ , \tag{c2}$$

$$\sigma^2 = \frac{1}{T}\left(\sum_{t=1}^{T} (y_t^1 - \mathbf{W}_t^1\alpha - \mathbf{X}_t^1\beta_1)^2 + \sum_{t=1}^{T} (y_t^2 - \mathbf{W}_t^2\alpha - \mathbf{X}_t^2\beta_2)^2 \right). \tag{c}$$

Man erhält für die in Kapitel 4.1 unterschiedenen Fälle:

[95] Vgl. Kapitel 2.2.2.

(1) $\mathbf{X} = \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Es sei

$$\overline{\mathbf{y}} = \begin{pmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \end{pmatrix} \text{ und } \overline{\mathbf{W}} = \begin{pmatrix} \mathbf{W}^1 \\ \mathbf{W}^2 \end{pmatrix}.$$

Die Normalgleichungen sind äquivalent zu denen des Modells

$$\overline{\mathbf{y}} = \overline{\mathbf{W}}\alpha + \mathbf{u}, \ \mathbf{u} \sim \mathrm{N}(0, \Omega), \ \Omega = \begin{pmatrix} \sigma_1^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I} \end{pmatrix}.$$

Dieser Fall wird in Kapitel 4.3.2.2.3 behandelt.

(2) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} = \mathbf{0}$, $\sigma_1^2 = \sigma_2^2$: Die Normalgleichungen sind (a1), (a2) und (c). Die Gleichungen (a1) bzw. (a2) sind die Normalgleichungen von separaten Regressionen von $\mathbf{y}^1$ auf $\mathbf{X}^1$ bzw. von $\mathbf{y}^2$ auf $\mathbf{X}^2$. Anders gesagt:

Sei
$$\overline{\mathbf{X}} = \begin{pmatrix} \mathbf{X}^1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^2 \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

und $\overline{\mathbf{y}}$ wie oben. Dann sind (a1) und (a2) die Normalgleichungen für β des gewöhnlichen linearen Modells

$$\overline{\mathbf{y}} = \overline{\mathbf{X}}\beta + \mathbf{u}, \ \mathbf{u} \sim \mathrm{N}(0, \sigma^2\mathbf{I}).$$

Also sind $\beta_1^{(n+1)}$ und $\beta_2^{(n+1)}$ die OLS-Schätzer einer Regression von $\overline{\mathbf{y}}$ auf $\overline{\mathbf{X}}$, d.h.

$$\begin{pmatrix} \beta_1^{(n+1)} \\ \beta_2^{(n+1)} \end{pmatrix} = \beta^{(n+1)} = (\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\overline{\mathbf{y}}.$$

Der gesuchte neue Parameterwert

$$\sigma^{2^{(n+1)}} = \frac{1}{T}(\overline{\mathbf{y}} - \overline{\mathbf{X}}\hat{\beta})'(\overline{\mathbf{y}} - \overline{\mathbf{X}}\hat{\beta}) = 2\left(\frac{1}{2T}(\overline{\mathbf{y}} - \overline{\mathbf{X}}\hat{\beta})'(\overline{\mathbf{y}} - \overline{\mathbf{X}}\hat{\beta})\right)$$

ist hier das Doppelte des ML-Schätzers für σ aus der Hilfsregression.

(3) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} = \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Die Normalgleichungen für β_1 und β_2 sind wieder (a1) und (a2). Also sind $\beta_1^{(n+1)}$ und $\beta_2^{(n+1)}$ dieselben wie bei **(2)**. Die Normalgleichungen für σ_1 und σ_2 sind (c1) und (c2). Es gilt

$$\sigma_i^{2^{(n+1)}} = \frac{1}{T}(\mathbf{y}^i - \mathbf{X}^i\beta_i^{(n+1)})'(\mathbf{y}^i - \mathbf{X}^i\beta_i^{(n+1)}), \ i = 1, 2.$$

(4) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} \neq \mathbf{0}$, $\sigma_1^2 = \sigma_2^2$: Die Normalgleichungen sind (a1), (a2), (b') und (c). Sei $\overline{\mathbf{y}}$ wie oben,

$$\overline{\mathbf{X}} = \begin{pmatrix} \mathbf{X}^1 & \mathbf{0} & \mathbf{W}^1 \\ \mathbf{0} & \mathbf{X}^2 & \mathbf{W}^2 \end{pmatrix} \text{ und } \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \alpha \end{pmatrix}.$$

Dann sind die Normalgleichungen für β die des Modells

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\beta + \mathbf{u}\,,\ \mathbf{u} \sim \mathrm{N}(0,\ \sigma^2\mathbf{I})\,,$$

d.h.

$$\begin{pmatrix} \beta_1^{(n+1)} \\ \beta_2^{(n+1)} \\ \alpha^{(n+1)} \end{pmatrix} = \beta^{(n+1)} = (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{y}}\ .$$

Für $\sigma^{2\,(n+1)}$ gilt

$$\sigma^{2\,(n+1)} = 2\left(\frac{1}{2T}(\bar{\mathbf{y}} - \bar{\mathbf{X}}\beta^{(n+1)})'(\bar{\mathbf{y}} - \bar{\mathbf{X}}\beta^{(n+1)})\right)\ .$$

(5) $\mathbf{X} \neq \mathbf{0}$, $\mathbf{W} \neq \mathbf{0}$, $\sigma_1^2 \neq \sigma_2^2$: Die Normalgleichungen sind (a1), (a2), (b), (c1) und (c2). Da (b) alle Parameter enthält, aber nicht linear ist, läßt sich hier kein geschlossener Ausdruck für die neuen Parameterwerte angeben. Sie sind die ML-Schätzer des Modells

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\beta + \mathbf{u}\,,\ \mathbf{u} \sim \mathrm{N}(\mathbf{0},\ \Omega),\ \Omega = \begin{pmatrix} \sigma_1^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I} \end{pmatrix}\ .$$

Dieser Fall wird im folgenden Kapitel behandelt.[96]

4.3.2.2.3 Berechnung der neuen Regressionsparameter und Störvarianz(en) für heteroskedastische Modelle mit regimeunabhängigen Parametern

Die Normalgleichungen des M-Schrittes für Fall **(1)** und für **(5)** sind die von linearen Modellen der Form

$$\mathbf{y} = \mathbf{Z}\gamma + \mathbf{u}\,,\ \mathbf{u} \sim \mathrm{N}(\mathbf{0},\ \Omega),\ \Omega = \begin{pmatrix} \sigma_1^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I} \end{pmatrix}\ .$$

Hier läßt sich wieder das Oberhofer-Kmenta-Verfahren aus Kapitel 4.2.1.3 verwenden (oder ein anderes numerisches Verfahren). KROLZIG [1997] schlägt in einer ähnlichen Situation vor, nicht $Q(\lambda \mid \mathbf{y}; \lambda^{(n)})$ zu maximieren, d.h. ML-Schätzer von Hilfsmodellen zu berechnen, sondern statt dessen die GLS-Schätzer für diese Hilfsmodelle zu verwenden (also nur den ersten Schritt des Oberhofer-Kmenta-Verfahrens durchzuführen).[97] Dieser Vorschlag führt auf einen sogenannten *Verallgemeinerten EM-Algorithmus* (*Generalised EM-Algorithm*) im Sinne von DEMPSTER, LAIRD, RUBIN [1977] und RUUD [1991]. Diese Autoren zeigen, daß die wesentlichen Eigenschaften des EM-Algorithmus (vgl. S. 54) bestehen bleiben, wenn im M-Schritt

[96] Für den Fall weiterer Regressoren $\mathbf{V}^1$ in Zustand 1 und $\mathbf{V}^2$ in Zustand 2 gilt wieder der Hinweis am Ende von Kapitel 4.2.1.2.

[97] Vgl. KROLZIG [1997], S. 109.

ein neuer Parameterwert gewählt wird, der den Wert der Likelihood im Vergleich zu dem aus dem letzten Iterationsschritt erhöht. Dies muß nicht notwendig der ML-Schätzer des Hilfsmodells sein. Dabei wird die Wahl eines neuen Parameters mit nur geringem Zuwachs der Likelihood die benötigte Anzahl von Iterationen des EM-Algorithmus erhöhen. In Fällen, in denen die Berechnung des M-Schrittes kompliziert ist, kann es jedoch möglich sein, so die Rechenzeit bis zur Konvergenz des EM-Algorithmus zu vermindern.

4.3.2.2.4 Berechnung der neuen Parameter des Zustandsprozesses

Für homogene Modelle lassen sich die Normalgleichungen des M-Schrittes für die Parameter des Zustandsprozesses nach den Parametern auflösen, so daß man geschlossene Ausdrücke für die neuen Parameterwerte erhält.

Für *Mischungsmodelle* gilt

$$Q(\lambda|\mathbf{y};\lambda^{(n)}) = \sum_{t=1}^{T}\left(\sum_{s_t}\left(\log f(y_t|s_t,\mathbf{X}_t,\mathbf{W}_t;\lambda)+\log P(s_t;\lambda)\right)P(s_t|\mathbf{y};\lambda^{(n)})\right)$$

mit $P(s_t=1;\lambda)=p$ und $P(s_t=2;\lambda)=1-p$. Durch Differentiation und Auflösen nach p erhält man

$$p^{(n+1)} = \frac{\sum_{t=1}^{T} P(s_t=1\,|\,\mathbf{y};\lambda^{(n)})}{T}.$$

Diese Formel hat eine sehr anschauliche Interpretation: Der neue Parameterwert für die Wahrscheinlichkeit dafür, daß Regime 1 herrscht, ist das arithmetische Mittel der geglätteten Wahrscheinlichkeiten für diesen Zustand. Wären die Zustände *beobachtbar*, so hätte $P(s_t=1\,|\,\mathbf{y};\lambda^{(n)})$ den Wert 0 oder 1, und $p^{(n+1)}$ wäre der gewöhnliche ML-Schätzer für die Erfolgswahrscheinlichkeit in einem Bernoulli-Experiment, berechnet aus einer einfachen Zufallsstichprobe vom Umfang T.

Für *Markov-Modelle* gilt

$$\begin{aligned} Q(\lambda\,|\,\mathbf{y};\lambda^{(n)}) = &\sum_{t=1}^{T}\left(\sum_{s_t}\log f(y_t\,|\,s_t,\mathbf{X}_t,\mathbf{W}_t;\lambda)P(s_t\,|\,\mathbf{y};\lambda^{(n)})\right) \\ &+\sum_{t=2}^{T}\left(\sum_{s_t,s_{t-1}}\log P(s_t\,|\,s_{t-1};\lambda)P(s_t,s_{t-1}\,|\,\mathbf{y};\lambda^{(n)})\right) \\ &+\sum_{s_1}\log P(s_1;\lambda)P(s_1\,|\,\mathbf{y};\lambda^{(n)}) \end{aligned}$$

mit $P(s_t=1\,|\,s_{t-1}=1;\lambda)=p^{11}$, $P(s_t=2\,|\,s_{t-1}=2;\lambda)=p^{22}$,

$P(s_t=2\,|\,s_{t-1}=1;\lambda)=1-p^{11}$, $P(s_t=1\,|\,s_{t-1}=2;\lambda)=1-p^{22}$,

$$P(s_1 = 1;\lambda) = \rho \text{ und } P(s_1 = 2;\lambda) = 1 - \rho .$$

Aus den Normalgleichungen für p^{11}, p^{22} und ρ erhält man

$$p_{11}^{(n+1)} = \frac{\sum_{t=2}^{T} P(s_t = 1, s_{t-1} = 1 \mid \mathbf{y};\lambda^{(n)})}{\sum_{t=2}^{T} P(s_{t-1} = 1 \mid \mathbf{y};\lambda^{(n)})}, \quad p_{22}^{(n+1)} = \frac{\sum_{t=2}^{T} P(s_t = 2, s_{t-1} = 2 \mid \mathbf{y};\lambda^{(n)})}{\sum_{t=2}^{T} P(s_{t-1} = 2 \mid \mathbf{y};\lambda^{(n)})}$$

und $$\rho^{(n+1)} = P(s_1 = 1 \mid \mathbf{y};\lambda^{(n)}) .$$

Diese Formeln lassen sich ähnlich anschaulich interpretieren wie die für Mischungsmodelle: Der neue Schätzer der Startwahrscheinlichkeit für Regime 1 ist die geglättete Wahrscheinlichkeit für diesen Zustand zum Zeitpunkt $t = 1$, und der neue Schätzer für p^{11}, die Wahrscheinlichkeit für ein Verbleiben in Regime 1, ist das arithmetische Mittel der geglätteten Wahrscheinlichkeiten für zweimaliges Aufeinanderfolgen dieses Zustandes, geteilt durch das arithmetische Mittel der geglätteten Wahrscheinlichkeiten dafür, daß der erste dieser Zustände Regime 1 war (und analog für p^{22}). Wären die Zustände *beobachtbar*, so wären dies die gewöhnlichen ML-Schätzer für die Übergangswahrscheinlichkeit einer homogenen Markov-Kette, berechnet aus einer Realisation der Kette von der Länge T.

Für *inhomogene* Modelle sind die Normalgleichungen für die Parameter des Zustandsprozesses in der Regel (das heißt für sinnvolle funktionale Formen für die Übergangs- bzw. die Aufenthaltswahrscheinlichkeiten) nicht linear und lassen sich nicht nach den Parametern auflösen. Diese Situation ist ähnlich wie in den Fällen, in denen sich die Normalgleichungen des M-Schrittes für die Regressionsparameter nicht exakt lösen lassen, vgl. Kapitel 4.3.2.2.3. DIEBOLD, LEE und WEINBACH [1994] haben (für inhomogene Markov-Modelle) vorgeschlagen, die Normalgleichungen durch eine Taylorentwicklung der Übergangswahrscheinlichkeiten bis zur ersten Ordnung zu linearisieren und diese approximativen Gleichungen nach den Parametern aufzulösen. In verschiedenen Testrechnungen zeigte sich, daß dieses Verfahren, das als verallgemeinerter EM-Algorithmus angesehen werden kann, ähnlich gute Eigenschaften hat wie der exakte EM-Algorithmus für homogene Modelle.

Die approximativen Gleichungen werden im Anhang B zu diesem Kapitel hergeleitet. Dabei wird zusätzlich zu den in DIEBOLD, LEE und WEINBACH [1994] untersuchten inhomogenen Markov-Modellen auch auf Mischungsmodelle und auf verschiedene Parameterrestriktionen eingegangen.

Anhang A: Berechnung der geglätteten Wahrscheinlichkeiten der Zustandsvariablen für Markov-Modelle

Alle folgenden Wahrscheinlichkeit(sdicht)en sind bedingt auf **X**, **W** und Y_0, berechnet mit den Parametern $\lambda^{(n)}$.

(I.) Zuerst werden die Größen

$$f(y_t|Y_{t-1})$$

und

$$P(s_t, s_{t-1}|Y_t)$$

iterativ für $t = 2,...,T$ berechnet, indem die folgenden beiden Relationen ausgenutzt werden:

(a)
$$\begin{aligned} f(y_t \mid Y_{t-1}) &= \sum_{s_t}\sum_{s_{t-1}}\sum_{s_{t-2}} f(y_t, s_t, s_{t-1}, s_{t-2} \mid Y_{t-1}) \\ &= \sum_{s_t}\sum_{s_{t-1}}\sum_{s_{t-2}} P(y_t \mid s_t) P(s_t \mid s_{t-1}) f(s_{t-1}, s_{t-2} \mid Y_{t-1}) \end{aligned}$$

und

(b)
$$\begin{aligned} P(s_t, s_{t-1} \mid Y_t) &= \frac{f(s_t, s_{t-1}, Y_t)}{f(Y_t)} \\ &= \frac{f(y_t \mid s_t) P(s_t \mid s_{t-1}) f(s_{t-1}, Y_{t-1})}{f(y_t \mid Y_{t-1}) f(Y_{t-1})} \\ &= \frac{f(y_t \mid s_t) P(s_t \mid s_{t-1}) P(s_{t-1} \mid Y_{t-1})}{f(y_t \mid Y_{t-1})} \\ &= \frac{f(y_t \mid s_t) P(s_t \mid s_{t-1}) \sum_{s_{t-2}} P(s_{t-1}, s_{t-2} \mid Y_{t-1})}{f(y_t \mid Y_{t-1})}. \end{aligned}$$

Gleichung (a) enthält außer den bekannten Größen ($f(y_t|s_t)$ und $P(s_t|s_{t-1})$) nur den Ausdruck, der in (b) berechnet wird, und (b) außer bekannten Größen nur den Ausdruck, der in (a) berechnet wird, sowie den, der im letzten Schritt von (b) berechnet wurde. Man kann für $t = 2,...,T$ die gesuchten Größen also iterativ ermitteln, indem man (a) und (b) abwechselnd berechnet. Für den Beginn der Iteration gilt

$$\begin{aligned} f(y_2 \mid y_1) &= \sum_{s_2}\sum_{s_1} f(y_2, s_2, s_1 \mid y_1) \\ &= \sum_{s_2}\sum_{s_1} f(y_2 \mid s_2) P(s_2 \mid s_1) P(s_1 \mid y_1). \end{aligned}$$

HAMILTON [1990] verwendet $P(s_1|y_1)$ als freien Parameter.[98] Für die Konstruktion der Log-Likelihood-Funktion wurde in dieser Arbeit hingegen wie in DIEBOLD, LEE und WEINBACH

[98] Der Grund dafür ist wohl, daß dadurch die sich aus dem M-Schritt ergebenden Formeln für die von ihm ausschließlich betrachteten Zeitreihenmodelle eine etwas intuitivere Gestalt bekommen.

[1994] als freier Parameter $P(s_1)$ verwendet,[99] deswegen muß $P(s_1|y_1)$ hier folgendermaßen berechnet werden:[100]

$$P(s_1|y_1) = \frac{f(s_1,y_1)}{f(y_1)} = \frac{f(y_1|s_1)P(s_1)}{\sum_{s_1} f(y_1|s_1)P(s_1)} .$$

Damit lautet der Iterationsanfang für (a) $f(y_2|y_1) = \dfrac{\sum_{s_2}\sum_{s_1} f(y_2|s_2)P(s_2|s_1)f(y_1|s_1)P(s_1)}{\sum_{s_1} f(y_1|s_1)P(s_1)}$

und für (b)

$$\begin{aligned} P(s_2,s_1|y_2,y_1) &= \frac{f(y_2,s_2,y_1,s_1)}{f(y_2,y_1)} \\ &= \frac{f(y_2|s_2)P(s_2|s_1)f(y_1|s_1)P(s_1)}{f(y_2|y_1)f(y_1)} \\ &= \frac{f(y_2|s_2)P(s_2|s_1)f(y_1|s_1)P(s_1)}{f(y_2|y_1)\sum_{s_1} f(y_1|s_1)P(s_1)} . \end{aligned}$$

(II.) Für ein festes $t \geq 2$ und $(s_t, s_{t-1}) \in \{1, 2\}\times\{1, 2\}$ berechnet man zuerst

$$P(s_{t+1},s_t,s_{t-1}|Y_{t+1}) = \frac{f(y_{t+1}|s_{t+1})P(s_{t+1}|s_t)P(s_t,s_{t-1}|Y_t)}{f(y_{t+1}|Y_t)}$$

und dann iterativ für $\tau = t+2,\ldots,T$

$$P(s_\tau,s_{\tau-1},s_t,s_{t-1}|Y_\tau) = \frac{\sum_{s_{\tau-2}} f(y_\tau|s_\tau)P(s_\tau|s_{\tau-1})P(s_{\tau-1},s_{\tau-2},s_t,s_{t-1}|Y_{\tau-1})}{f(y_\tau|Y_{\tau-1})} .$$

(III.) Wenn man $\tau = T$ erreicht hat, berechnen sich daraus die geglätteten gemeinsamen Wahrscheinlichkeiten aufeinanderfolgender Zustandsvariablen folgendermaßen:

$$P(s_t,s_{t-1}|Y_T) = \sum_{s_T}\sum_{s_{T-1}} P(s_T,s_{T-1},s_t,s_{t-1}|Y_T) .$$

(IV.) Daraus erhält man schließlich die geglätteten Wahrscheinlichkeiten der Zustandsvariablen vermöge

$$P(s_t|Y_T) = \sum_{s_{t-1}} P(s_t,s_{t-1}|Y_T) .$$

[99] Daß diese Autoren $P(s_1)$ als freien Parameter verwenden, erkennt man an ihrer Formel für die *complete-data likelihood* auf S. 286.

[100] Dies haben DIEBOLD, LEE und WEINBACH [1994] in ihrer ersten Formel auf S. 290 nicht beachtet. Ihr Verfahren zur Berechnung der geglätteten Wahrscheinlichkeiten (und damit der E-Schritt ihres EM-Algorithmus) ist deshalb nicht korrekt.

Anhang B: Berechnung der neuen Parameterschätzer des Zustandsprozesses für inhomogene Mischungs- und Markov-Modelle

Für *Mischungsmodelle* gilt[101]

$$Q(\lambda \mid \mathbf{y};\lambda^{(n)}) = \sum_{t=1}^{T}\left(\sum_{s_t}\left(\log f(y_t \mid s_t, \mathbf{X}_t, \mathbf{W}_t;\lambda) + \log P(s_t \mid z_t;\lambda)\right)P(s_t \mid \mathbf{y};\lambda^{(n)})\right).$$

Sei

$$g_t = g(z_t\gamma + \delta) = P(s_t = 1 \mid z_t;\lambda)$$

und

$$g_t' = \left.\frac{dg(z)}{dz}\right|_{z=z_t\gamma+\delta} .$$

Dann lautet die Normalgleichungen für δ

$$\sum_{t=1}^{T}\frac{g_t'}{g_t(1-g_t)}\left(P(s_t = 1 \mid \mathbf{y};\lambda^{(n)})(1-g_t) + (1-P(s_t = 1 \mid \mathbf{y};\lambda^{(n)}))g_t\right) = 0$$

und die für γ

$$\sum_{t=1}^{T}\frac{z_t g_t'}{g_t(1-g_t)}\left(P(s_t = 1 \mid \mathbf{y};\lambda^{(n)})(1-g_t) + (1-P(s_t = 1 \mid \mathbf{y};\lambda^{(n)}))g_t\right) = 0 .$$

Wählt man für g_t die logistische Funktion, so gilt[102]

$$\frac{g_t'}{g_t(1-g_t)} = 1,$$

so daß sich obige Gleichungen vereinfachen zu

$$\sum_{t=1}^{T} P(s_t = 1 \mid \mathbf{y};\lambda^{(n)}) = \sum_{t=1}^{T} g_t$$

und

$$\sum_{t=1}^{T} z_t P(s_t = 1 \mid \mathbf{y};\lambda^{(n)}) = \sum_{t=1}^{T} z_t g_t .$$

Ersetzt man nun g_t durch die Taylorentwicklung bis zur ersten Ordnung um den letzten Parameterwert $\lambda^{(n)}$:

$$\begin{aligned} g_t &\approx g(z_t\gamma^{(n)} + \delta^{(n)}) + \left.\frac{\partial g}{\partial \delta}\right|_{z=z_t\gamma^{(n)}+\delta^{(n)}} \cdot (\delta - \delta^{(n)}) + \left.\frac{\partial g}{\partial \gamma}\right|_{z=z_t\gamma^{(n)}+\delta^{(n)}} \cdot (\gamma - \gamma^{(n)}) \\ &= g(z_t\gamma^{(n)} + \delta^{(n)}) + \left.\frac{dg}{dz}\right|_{z=z_t\gamma^{(n)}+\delta^{(n)}} \cdot \left((\delta - \delta^{(n)}) + z_t(\gamma - \gamma^{(n)})\right), \end{aligned}$$

so erhält man lineare Gleichungen in δ und γ. Mit den Bezeichnungen

[101] Im folgenden wird nur der für empirische Arbeiten bei weitem wichtigste Fall eines skalaren Indikators z_t betrachtet. Verallgemeinerungen für inhomogene Markov-Modelle mit vektoriellen Indikatoren finden sich in DIEBOLD, LEE und WEINBACH [1994], S. 300 f.

[102] Diese Beziehung macht die Verwendung der logistischen Funktion vorteilhaft.

$$g_t^{(n)} = g(z_t\gamma^{(n)} + \delta^{(n)})$$

und
$$g_t^{(n)\prime} = \frac{dg(z)}{dz}\Big|_{z=z_t\gamma^{(n)}+\delta^{(n)}}$$

lautet die Lösung dieser Gleichungen

$$\begin{pmatrix}\delta^{(n+1)}\\ \gamma^{(n+1)}\end{pmatrix} = \begin{pmatrix}\sum_{t=1}^{T} g_t^{(n)\prime} & \sum_{t=1}^{T} z_t g_t^{(n)\prime}\\ \sum_{t=1}^{T} z_t g_t^{(n)\prime} & \sum_{t=1}^{T} z_t^2 g_t^{(n)\prime}\end{pmatrix}^{-1} \cdot$$

$$\begin{pmatrix}\sum_{t=1}^{T}\left(P(s_t = 1 \mid y;\lambda^{(n)}) - \left[g_t^{(n)} - g_t^{(n)\prime}(\delta^{(n)} + z_t\gamma^{(n)})\right]\right)\\ \sum_{t=1}^{T} z_t\left(P(s_t = 1 \mid y;\lambda^{(n)}) - \left[g_t^{(n)} - g_t^{(n)\prime}(\delta^{(n)} + z_t\gamma^{(n)})\right]\right)\end{pmatrix}.$$

Falls die Restriktion $\delta = 0$ verwendet wird,[103] so gilt

$$\gamma^{(n+1)} = \frac{\sum_{t=1}^{T} z_t\left(P(s_t = 1 \mid y;\lambda^{(n)}) - \left[g_t^{(n)} - g_t^{(n)\prime}(\delta^{(n)} + z_t\gamma^{(n)})\right]\right)}{\sum_{t=1}^{T} z_t^2 g_t^{(n)\prime}}.$$

Die Herleitung der Näherungsformel für *Markov-Modelle* ist völlig analog. Sie findet sich in DIEBOLD, LEE und WEINBACH [1994], S. 300 f., und wird deshalb an dieser Stelle nur verkürzt wiedergegeben. Mit der Bezeichnung $P(s_1 = 1;\lambda) = \rho$ erhält man aus der Normalgleichung für ρ wieder

$$\rho^{(n+1)} = P(s_1 = 1 \mid \mathbf{y};\lambda^{(n)}) \ .$$

Sei
$$g_{i,t} = g(z_t\gamma_i + \delta_{i1}) = P(s_t = i \mid s_t = i, z_t;\lambda),$$

$$g_{i,t}{}' = \frac{dg(z)}{dz}\Big|_{z=z_t\gamma_i+\delta_i},$$

$$g_{i,t}^{(n)} = g(z_t\gamma_i^{(n)} + \delta_i^{(n)})$$

und
$$g_{i,t}^{(n)\prime} = \frac{dg(z)}{dz}\Big|_{z=z_t\gamma_i^{(n)}+\delta_i^{(n)}} \quad \text{für } i = 1, 2.$$

[103] Dies ist der Fall eines bekannten "Umschlagpunktes" für die Aufenthaltswahrscheinlichkeiten bei $z_t = 0$. Für einen anderen bekannten Umschlagpunkt δ kann man obige Formel verwenden, wenn man z_t durch $z_t - \delta$ ersetzt.

Die Näherungsformel lautet

$$\begin{pmatrix} \delta_i^{(n+1)} \\ \gamma_i^{(n+1)} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T} P(s_{t-1} = i \mid y; \beta^{(n)}) g_{i,t}^{(n)\prime} & \sum_{t=1}^{T} z_t P(s_{t-1} = i \mid y; \beta^{(n)}) g_{i,t}^{(n)\prime} \\ \sum_{t=1}^{T} z_t P(s_{t-1} = i \mid y; \beta^{(n)}) g_{i,t}^{(n)\prime} & \sum_{t=1}^{T} z_t^2 P(s_{t-1} = i \mid y; \beta^{(n)}) g_{i,t}^{(n)\prime} \end{pmatrix}^{-1} \cdot$$

$$\begin{pmatrix} \sum_{t=1}^{T} \begin{pmatrix} P(s_t = i, s_{t-1} = i \mid y; \beta^{(n)}) \\ - P(s_{t-1} = i \mid y; \beta^{(n)}) \left[g_{i,t}^{(n)} - g_{i,t}^{(n)\prime} (\delta_i^{(n)} + z_t \gamma_i^{(n)}) \right] \end{pmatrix} \\ \sum_{t=1}^{T} z_t \begin{pmatrix} P(s_t = i, s_{t-1} = i \mid y; \beta^{(n)}) \\ - P(s_{t-1} = i \mid y; \beta^{(n)}) \left[g_{i,t}^{(n)} - g_{i,t}^{(n)\prime} (\delta_i^{(n)} + z_t \gamma_i^{(n)}) \right] \end{pmatrix} \end{pmatrix}$$

für $i = 1, 2$. Falls die Restriktion $\delta = 0$ verwendet wird,[104] so gilt

$$\gamma^{(n+1)} = \frac{\sum_{t=1}^{T} z_t \begin{pmatrix} P(s_t = i, s_{t-1} = i \mid y; \lambda^{(n)}) \\ - P(s_{t-1} = i \mid y; \lambda^{(n)}) \left[g_{i,t}^{(n)} - g_{i,t}^{(n)\prime} (\delta_i^{(n)} + z_t \gamma_i^{(n)}) \right] \end{pmatrix}}{\sum_{t=1}^{T} z_t^2 P(s_{t-1} = i \mid y; \lambda^{(n)}) g_{i,t}^{(n)\prime}} .$$

[104] Hier gilt sinngemäß wieder Fußnote 103.

5 ML-Schätzung von Regimewechselmodellen für Wochendaten des FF/DM-Wechselkurses

Die Schätzverfahren für Regimewechselmodelle werden im folgenden an der Modellierung von wöchentlichen Durchschnittswerten des Wechselkurses zwischen dem Französischem Franc (FF) und der Deutschen Mark (DM) veranschaulicht.

Der Beobachtungszeitraum wurde so gewählt, daß er keine Realignments enthält, da mit Realignments Wechselkurssprünge verbunden sind, die exogener Natur sind und die Schätzung wesentlich verzerren würden, wenn ihnen in der Modellierung nicht separat Rechnung getragen wird. Da das letzte Realignment, das den FF/DM-Kurs betraf, am 12. Januar 1987 stattfand, wird für die Untersuchung der Zeitraum vom 23. Januar 1987 bis zum 20. November 1998 zugrunde gelegt.[105] Damit umfaßt die Reihe (mit insgesamt 618 Werten) praktisch den gesamten Zeitraum zwischen dem letzten FF/DM-Realignment und dem Beginn der Europäischen Währungsunion, also dem Ende des EWS.

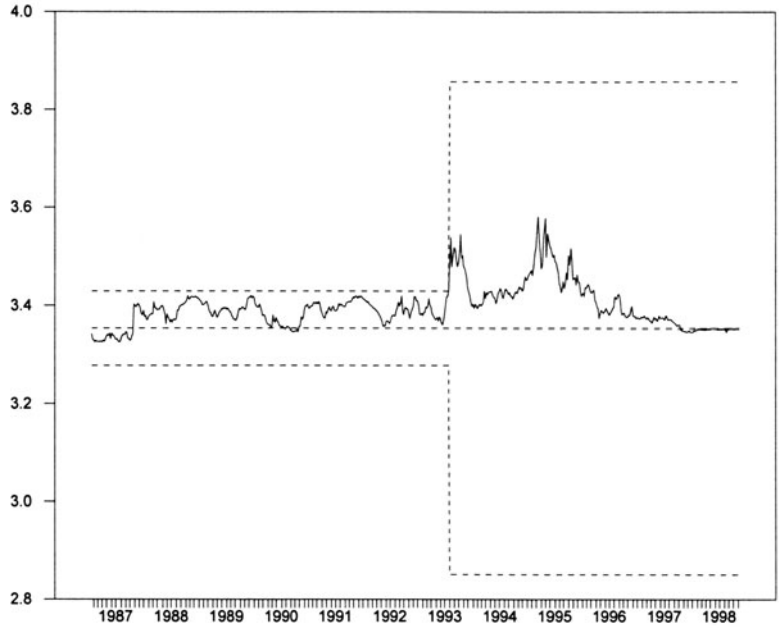

Abbildung 4: Wechselkurs zwischen FF und DM, Leitkurs und Interventionsgrenzen

[105] Quelle: Datastream.

Es bezeichne w_t den Preis einer DM in FF. Der Verlauf von w_t ist zusammen mit dem FF/DM-Leitkurs und den zugehörigen Interventionsgrenzen im EWS in Abbildung 4 wiedergegeben. Deutlich zu sehen ist die Erweiterung der Bandbreiten von 4,5 % auf 30 % am 2. August 1993. Dies legt die Frage nahe, ob in der Beobachtungsperiode ein Strukturbruch stattgefunden hat. Aufgrund der Flexibilität der verwendeten Modelle ist es jedoch nicht nötig, solch einen eventuellen Strukturbruch separat zu modellieren. Er müßte sich in einem Mischungs- oder Markov-Modell in einem abrupten Umschlag der geglätteten Wahrscheinlichkeiten an diesem Zeitpunkt niederschlagen.

Der Wechselkurs ist in der betrachteten Periode wiederholt in die Nähe des oberen, nie jedoch des unteren Interventionspunktes gelangt. Dies läßt die Vermutung zu, daß Phasen mit „normalem" Wechselkursverlauf abwechselten mit solchen, in denen darauf spekuliert wurde, daß der Franc gegenüber der Mark abgewertet würde (wie es bis Anfang 1987 schon sechs mal geschehen war). Ein Modell mit zwei Zuständen erscheint hier also gerechtfertigt.

Den Gegenstand der empirischen Untersuchung bilden die wöchentlichen Differenzen der logarithmierten und mit 100 multiplizierten Werte des Wechselkurses, also

$$r_t = 100(\log w_t - \log w_{t-1}) = 100 \log(w_t / w_{t-1}) \ .$$

Für die Wahl dieser Differenzen lassen sich verschiedene Gründe anführen:

- Die Differenzen können als stationär angesehen werden (s.u.). Deswegen werden die Schätzer für Regimewechselmodelle gute Eigenschaften für große Stichproben haben, vgl. Kapitel 6.1.
- Diese logarithmischen Differenzen sind (im Gegensatz zu Veränderungsraten, die den Wert –1 nicht unterschreiten können) nicht nach unten begrenzt. Für sie ist also eine symmetrische Verteilung möglich.
- Das Standard-Optionspreismodell von Black und Scholes[106] geht davon aus, daß diese Größen unabhängig identisch normalverteilt sind. Diese Annahme (die auch an anderen Stellen in der Finanzmarktökonometrie getroffen wird) kann direkt mit den Untersuchungsergebnissen für die beiden betrachteten Wechselkurse kontrastiert werden.

In Tabelle 1 sind einige wichtige statistische Daten der Verteilung von r_t wiedergegeben. Auffällig ist der hohe Wert der Kurtosis. Die Nullhypothese einer Kurtosis von Null, wie es für die Normalverteilung der Fall ist, wird deutlich verworfen. Die Schiefe dagegen ist mode-

[106] Vgl. z.B. JARROW und RUDD [1983].

rat und nicht signifikant von Null verschieden.[107] Die hohe Kurtosis ist auch der Grund dafür, daß der Jarque-Bera-Test die Nullhypothese einer Normalverteilung verwirft.[108]

Größe / Teststatistik	Wert	Signifikanzniveau
Minimum	-2,254	
Maximum	1,795	
Maximaler Wert der Log-Likelihood	-169,36	
Mittelwert μ	0,00058	
Standardabweichung σ	0,319	
Schiefe	0,056	0,57
Kurtosis	11,52	0,00
Jarque-Bera	3400,77	0,00
Durbin-Watson	2,13	>0,05
LB-10 / 20 (für r_t)[a]	7,15 / 20,92	0,71 / 0,40[b]
Dickey-Fuller	-11,50	<0,01
LB-10 / 20 (für r_t^2)[a]	252,06 / 265,96	0,00 / 0,00[c]
ARCH(1)	120,13	0,00

Tabelle 1: Statistische Daten der Verteilung von r_t
[a]: LB-10 / 20: Ljung-Box-Test bis zum 10. / 20. Lag
[b]: Zum 5 %-Niveau signifikant: 11. Lag
[c]: Zum 5 %-Niveau signifikant: 1. bis 11. Lag

Der Wert der Durbin-Watson-Statistik ist nicht signifikant zum 5 %-Niveau. Der Ljung-Box-Test enthüllt keine signifikante Autokorrelation in den ersten 10 bzw. 20 Lags. In der Bartlett-Approximation ist nur ein einziger Lag (der 11.) zum 5 %-Niveau signifikant, was bei 20 geprüften empirischen Autokorrelationen gerade das erwartete Ergebnis unter der Nullhypothese nicht vorhandener Autokorrelation ist.

Da es sich um differenzierte Daten handelt, verwundert es nicht, daß der Dickey-Fuller-Test die Nullhypothese eines Random-Walks zugunsten eines AR(1)-Modells verwirft, wobei angesichts der niedrigen empirischen Autokorrelation davon auszugehen wäre, daß der AR(1)-Parameter gleich Null ist.

Deutlich sind auch die Ergebnisse der Ljung-Box-Tests für die quadrierten Beobachtungen, die zumindest unter der Nullhypothese von weißem Rauschen ohne Modifikationen angewendet werden können.[109] Die Nullhypothese wird deutlich verworfen, und die ersten 11 unter den ersten 20 empirischen Autokorrelationen sind signifikant von Null verschieden. Dies

[107] Zur Verteilung der empirischen Schiefe und Kurtosis unter der Nullhypothese einer Normalverteilung vgl. z.B. SNEDECOR und COCHRANE [1980], S. 552.

[108] Vgl. JARQUE und BERA [1982].

[109] Vgl. MCLEOD und LI [1983].

zeigt sich auch an dem signifikanten Ergebnis des Lagrange-Multiplikatoren-Tests der Nullhypothese von weißem Rauschen gegen die Alternative eines ARCH(1)-Modells.[110]

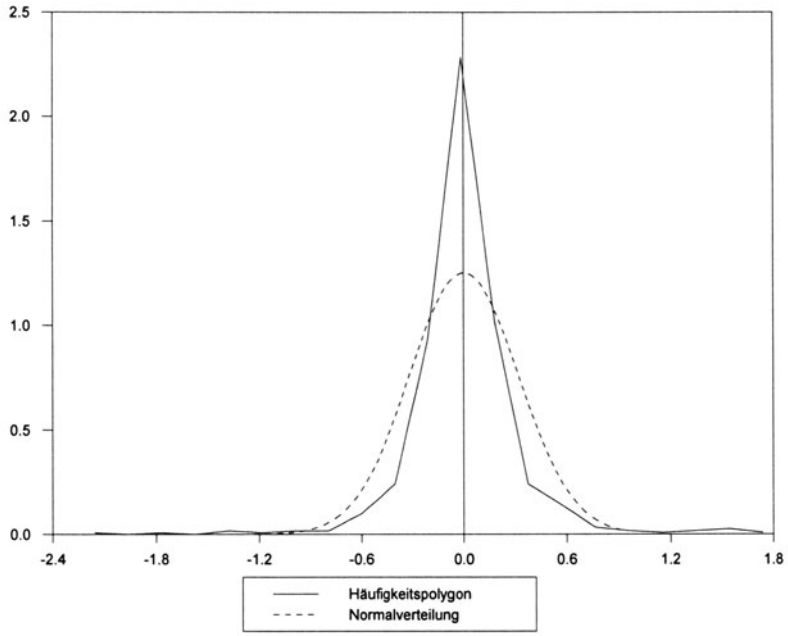

Abbildung 5: Häufigkeitspolygon und Dichte der geschätzten Normalverteilung von r_t

Die wichtigsten *stylised facts* der Daten sind also

- die starke Leptokurtosis, die sich auch darin zeigt, daß das empirische Häufigkeitspolygon (oder ein anderer Dichteschätzer) deutlich höhere Werte in der Nähe des Mittelwertes und über den *tails* aufweist als die entsprechende Normalverteilung (vgl. Abbildung 5),
- die hohe Autokorrelation der quadrierten Werte, die auf das gehäufte Auftreten von Werten mit hoher bzw. niedriger Volatilität (*volatility clustering*) schließen läßt.

Eine Möglichkeit, solche Daten zu modellieren, stellt ein ARCH-Prozeß mit leptokurtischer Verteilung der Störterme dar. Im folgenden soll aber gezeigt werden, wie sich diese Charakteristika der Zeitreihe mit Hilfe von Regimewechselmodellen beschreiben und erklären lassen.

[110] Vgl. BOLLERSLEV, ENGLE und NELSON [1994].

5.1 Ein Schwellenmodell

Bei Gültigkeit der ungedeckten Zinsparität entspricht die Zinsdifferenz zwischen zwei Ländern der erwarteten Veränderungsrate des Wechselkurses der betroffenen Währungen. Niedrige Zinsdifferenzen sind also ein Indikator für geringe erwartete Abwertungen der Währung mit dem höheren Zins, die sich dann noch innerhalb des Wechselkursbandes vollziehen können, hohe Zinsdifferenzen sind ein Anzeichen dafür, daß damit zu rechnen ist, daß diese Währung durchaus auch das Wechselkursband verlassen kann, d.h. daß ein Realignment stattfinden muß. In einer solchen Situation kann Spekulation gegen die abwertungsgefährdete Währung profitabel werden. Es ist deshalb vorgeschlagen worden, die Zinsdifferenz als Indikator für einen Regimewechsel zu verwenden:[111] Überschreitet sie eine gewisse kritische Schwelle, so setzt sich am Markt die Überzeugung durch, daß die gegenwärtigen Leitkurse nicht aufrechterhalten werden können, und es beginnt eine spekulative Phase.

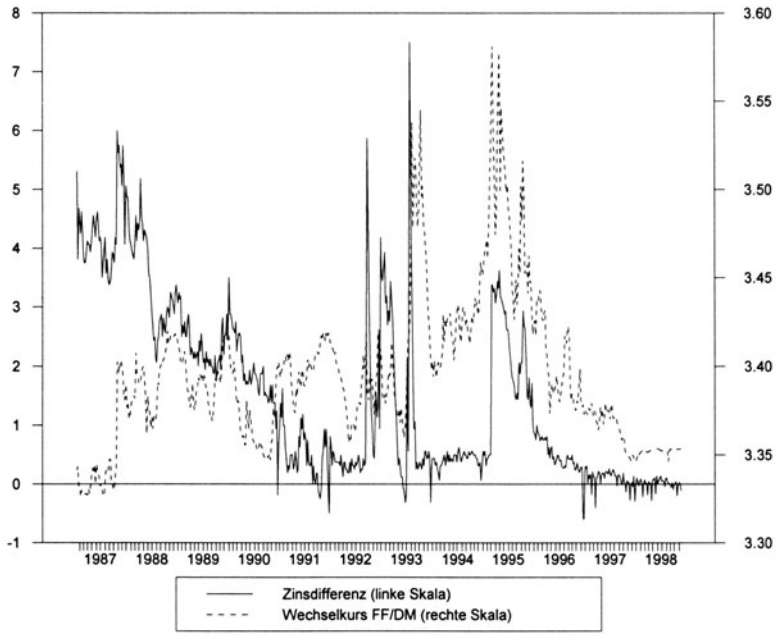

Abbildung 6: Differenz zwischen FF- und DM-Zinsen und FF/DM-Wechselkurs

[111] Vgl. BEKAERT und GRAY [1998].

In Abbildung 6 ist die Entwicklung der Differenz zwischen Zinsen für FF- und für DM-Einlagen dem Verlauf des Wechselkurses gegenübergestellt.[112]

Bis etwa 1990 waren die Zinsen für FF-Einlagen deutlich höher als für DM-Einlagen (mit sinkender Tendenz), es wurde also eine Risikoprämie für FF-Einlagen verlangt. Ab Ende 1990 bis zur Krise des EWS 1992/93, die sich in hohen Ausschlägen der Zinsdifferenz zeigte, ging die Zinsdifferenz zurück. Nach einem Rückgang auf das Niveau vor der Krise stieg die Zinsdifferenz 1995 erneut stark an, um danach wieder zurückzugehen und Ende 1998, also unmittelbar vor Beginn der Währungsunion, in den Wert Null zu münden, worin sich das Vertrauen in das Übergehen des EWS-Leitkurses in die FF/DM-Relation in der Währungsunion widerspiegelte.

Das hier untersuchte Modell ist demnach ein Schwellenmodell mit der Zinsdifferenz als Schwellenvariable und zu schätzendem Schwellenwert. Die entsprechenden Schätzwerte sind in Tabelle 2 dargestellt.[113]

$\hat{c}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,25	0,0031 (0,010)	$-9{,}36 \cdot 10^{-5}$ (0,016)	0,116	0,352	-90,61

Tabelle 2: Geschätzte Parameter eines Schwellenmodells für r_t

Die beiden Zustände unterscheiden sich deutlich in den Volatilitäten. Die Volatilität ist höher in dem Zustand, der durch Zinsdifferenzen oberhalb des geschätzten Schwellenwertes von 0,25 % charakterisiert ist. Diese Beobachtung steht im Einklang mit der Hypothese, daß hohe Zinsdifferenzen einen Indikator für spekulatives Verhalten darstellen.[114]

Der geschätzte Schwellenwert erscheint angesichts der üblichen Zinsdifferenzen als sehr niedrig. Teilt man die Beobachtungsperiode in Phasen auf, die zu den beiden Zuständen entsprechend diesem geschätzten Schwellenwert gehören, so überwiegt Zustand 1 mit der hohen Volatilität (vgl. Abbildung 7).

Bis 1997 wird hauptsächlich dieser Zustand beobachtet (mit einer größeren Zahl von kurzen Unterbrechungen), und danach (mit einer kurzen Unterbrechung) ausschließlich der andere. Das Schwellenmodell hat also beinahe den Charakter eines Strukturbruchmodells, mit

[112] Es handelt sich um die Differenzen zwischen auf Jahresbasis umgerechnete Zinsen für einwöchige FF- und DM-Einlagen. Da der französische Kapitalmarkt erst im Verlaufe der achtziger Jahre vollständig liberalisiert wurde, spiegeln an Märkten in Frankreich erhobene Zinsen die Marktbedingungen u.U. nicht vollständig wider. Deswegen wurden Zinsen für Einlagen in London verwendet (*eurocurrency rates*). Quelle: Datastream.

[113] Zur Berechnung der ML-Schätzer vgl. den Anhang zu diesem Kapitel. - In Klammern sind die geschätzten Standardfehler der Schätzwerte angegeben.

[114] Auf die beiden Mittelwerte wird in Kapitel 7.3 genauer eingegangen.

einem Strukturbruch, der durch einen permanenten Übergang zu niedrigeren Volatilitäten ab etwa Anfang 1997 charakterisiert ist.

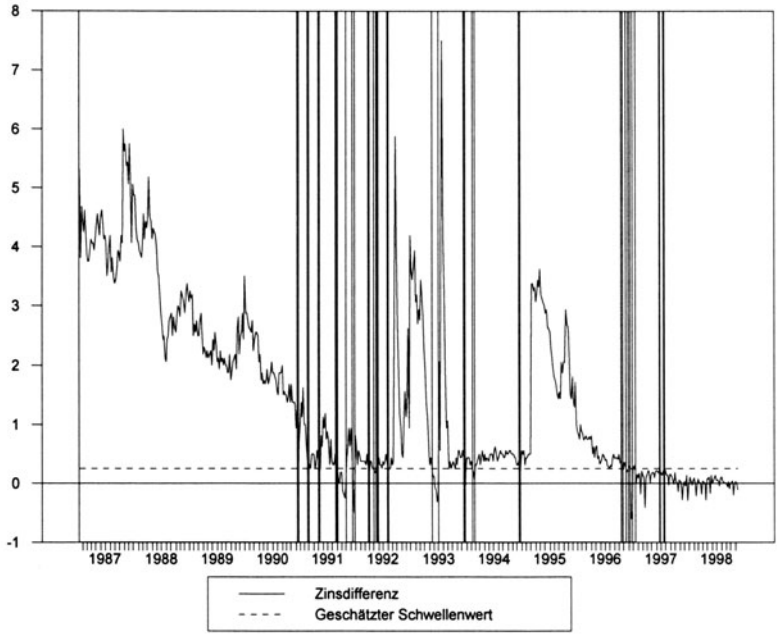

Abbildung 7: Abfolge der Zustände im Schwellenmodell (senkrechte Linien entsprechen Zustandswechseln)

Zumindest qualitativ kann dieses Modell die oben erwähnten *stylised facts* der Daten erklären: Es treten lange Phasen hoher und niedriger Volatilität auf, die für die beobachtete Autokorrelation der quadrierten Werte verantwortlich sind. Ihre Mischung erklärt die hohe empirische Leptokurtosis. Diese Aspekte können aber auch durch andere Modelle beschrieben werden, wie die folgenden Untersuchungen zeigen.

5.2 Ein Mischungsmodell

In Tabelle 3 sind die Schätzwerte eines Mischungsmodells für die Zeitreihe r_t nach dem ML-Verfahren angegeben.[115]

[115] In Klammern sind die mit einer geschätzten Varianz-Kovarianz-Matrix berechneten Standardfehler der Schätzwerte angegeben. Da man mit diesem Verfahren auch geschätzte Standardfehler der geschätzten Standardabweichungen der Störterme erhält, sind diese der Vollständigkeit halber in der Tabelle aufgeführt. Vgl. hierzu und zur Berechnung der ML-Schätzer den Anhang zu diesem Kapitel.

$\hat{p}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,189 (0,027)	0,060 (0,065)	-0,013 (0,0077)	0,661 (0,041)	0,148 (0,0064)	-3,37

Tabelle 3: Geschätzte Parameter eines Mischungsmodells für r_t

Charakteristisch sind die um einen Faktor von mehr als 4 verschiedenen Standardabweichungen. In diesem Modell tritt also mit einer geschätzten Wahrscheinlichkeit von 0,189 ein Zustand mit hoher Volatilität ein. Die geschätzten Standardfehler der Schätzwerte für p, σ_1 und σ_2 sind relativ klein und die für μ_1 und μ_2 verhältnismäßig groß.[116]

Aus den geschätzten Parametern lassen sich die Momente der geschätzten Mischungsverteilung und ihre Dichte berechnen.[117] Die Momente sind in Tabelle 4 wiedergegeben. Ein Vergleich mit Tabelle 1 zeigt, daß sie nahe bei den empirischen Momenten liegen.[118]

Mittelwert	Standardabweichung	Schiefe	Kurtosis
0,00058	0,318	0,43	10,84

Tabelle 4: Momente der geschätzten Mischungsverteilung

In Abbildung 8 wird die Dichtefunktion der Mischungsverteilung dem empirischen Häufigkeitspolygon (und der geschätzten Normalverteilung) gegenübergestellt. Es zeigt sich, daß die Dichte der Mischungsverteilung die empirische Häufigkeitsverteilung gut beschreibt.

Zusammenfassend läßt sich sagen, daß Mischungsverteilungen in der Lage sind, ein zentrales Charakteristikum der untersuchten Daten, ihre auffällig hohe Kurtosis, zu beschreiben und zu erklären: Es treten abwechselnd Phasen hoher und niedriger Volatilität auf, deren Mischung zu der beobachteten Leptokurtosis führt.

Rückschlüsse auf die Abfolge der beiden Zustände in den Daten lassen sich mit Hilfe der geglätteten Wahrscheinlichkeiten ziehen. In Abbildung 9 ist der zeitliche Verlauf dieser Wahrscheinlichkeit für Zustand 1 (mit hoher Volatilität) dargestellt.

Da die Zustände in einem Mischungsmodell unabhängig sind, ist die Kurve sehr erratisch. Dennoch lassen sich einige charakteristische Phasen ausmachen. Bis Anfang 1992 ist die Wahrscheinlichkeit für den volatilen Zustand insgesamt niedrig (von einzelnen Ausreißern abgesehen). Markant ist insbesondere das letzte Jahr dieser Periode, in dem weitgehendes Vertrauen der Märkte in das EWS eingekehrt zu sein scheint.

---

[116] Auf die beiden Mittelwerte wird in Kapitel 7.3 genauer eingegangen.

[117] Zur Berechnung dieser Momente vgl. KAEHLER und MARNET [1994], S. 212.

[118] D.h. daß Schätzwerte nach der Methode der Momente vermutlich nicht wesentlich von diesen ML-Schätzwerten abweichen.

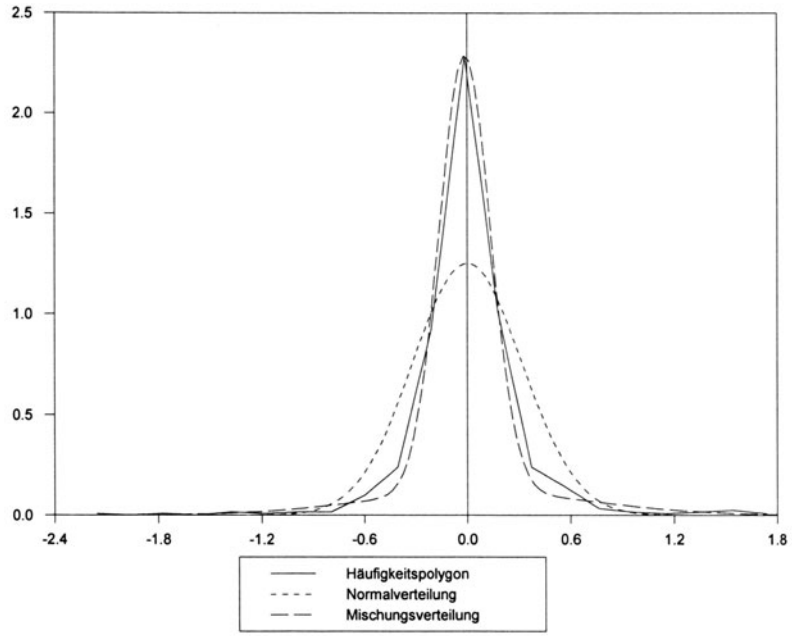

Abbildung 8: Häufigkeitspolygon, Dichte der geschätzten Normal- und Mischungsverteilung von r_t

Ab Mitte 1992 zeigt sich die Krise des EWS in höheren Wahrscheinlichkeiten für den volatilen Zustand, die bis etwa Ende 1996 andauern. Ein eventuell vorhandener Strukturbruch nach der Erweiterung der Bandbreiten im August 1993 läßt sich an den geglätteten Wahrscheinlichkeiten nicht erkennen.

Die niedrigen geglätteten Wahrscheinlichkeiten für Zustand 1 ab Ende 1996 zeigen, daß sich zu diesem Zeitpunkt die Überzeugung durchsetzte, daß der seit Januar 1987 bestehende Leitkurs marktgerecht war und auch die FF/DM-Relation nach Inkrafttreten der Europäischen Währungsunion am 1. Januar 1999 darstellen würde, wie es dann tatsächlich eingetreten ist.

Auffällig ist, daß die geglätteten Wahrscheinlichkeiten ein Minimum von etwa 0,05 nicht unterschreiten. Dies ist kein Zufall, sondern beruht auf der prinzipiellen Schwäche von Mischungsmodellen, einen Zustand mit niedriger Volatilität sicher zu erkennen, wenn die Mittelwerte der Zustände nahe beieinander liegen (wie es hier der Fall ist). An der Formel für die geglättete Wahrscheinlichkeit aus Kapitel 4.3.2.1 (S. 57) erkennt man, daß sie nicht beliebig klein werden kann, weil auch für Beobachtungen in der Nähe des gemeinsamen Mittelwertes die Dichte der Komponente mit der hohen Volatilität nicht Null wird. Anders ist dies

bei absolut gesehen hohen Beobachtungen, bei denen die Dichte der Komponente mit der niedrigen Volatilität gegen Null geht und deshalb die geglättete Wahrscheinlichkeit für Zustand mit der hohen Volatilität gegen Eins.

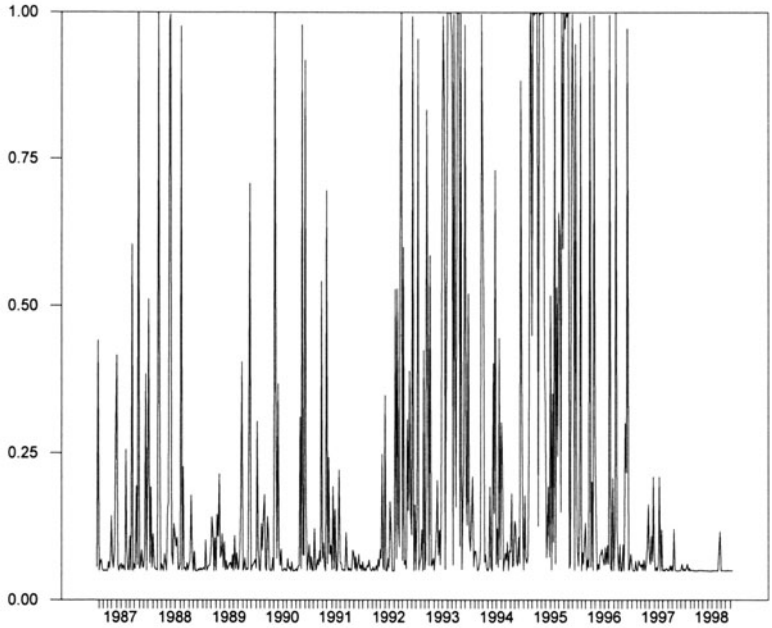

Abbildung 9: Geglättete Wahrscheinlichkeit für Zustand 1 (Mischungsmodell)

5.3 Ein Markov-Modell

Ein Mischungsmodell (bei dem die Beobachtungen unabhängig voneinander sind) kann die hohe Autokorrelation der quadrierten Beobachtungen nicht erklären. Sie hat ihren Grund darin, daß in den Daten Beobachtungen mit großen positiven oder negativen Ausschlägen, also mit hoher Volatilität, gehäuft auftreten (und ebenso Beobachtungen mit niedriger Volatilität). Es liegt also bedingte Heteroskedastizität vor. In der Sprache von Regimewechselmodellen: Die Wahrscheinlichkeit ist hoch, daß auf einen Zustand mit hoher Volatilität ein ebensolcher folgt, und ebenso für den anderen Zustand. Die Unabhängigkeit der Zustände, die für ein Mischungsmodell charakteristisch ist, liegt hier also vermutlich nicht vor. Es bietet sich an, statt dessen ein Markov-Modell zu verwenden, in dem die Wahrscheinlichkeit für einen

Zustand explizit vom Zustand der Vorperiode abhängt. Die Schätzergebnisse für ein solches Modell sind in Tabelle 5 dargestellt.[119]

$\hat{p}^{11}$	$\hat{p}^{22}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,839 (0,052)	0,952 (0,016)	0,030 (0,055)	-0,0082 (0,0071)	0,611 (0,085)	0,142 (0,013)	44,12

Tabelle 5: Geschätzte Parameter eines Markov-Modells für r_t

Die Schätzwerte für die Mittelwerte und die Standardabweichungen (und sogar die geschätzten Standardfehler der Schätzwerte) sind denen des Mischungsmodells (vgl. Tabelle 3) ähnlich. Auffällig sind die hohen Schätzwerte der Übergangswahrscheinlichkeiten $\hat{p}^{11}$ und $\hat{p}^{22}$, die die Wahrscheinlichkeiten für das Verbleiben in den beiden Zuständen darstellen. Der Grund hierfür liegt in der Häufung von Beobachtungen aus jeweils einem Zustand. Der sich daraus ergebende Erwartungswert der Verweildauer in diesen Zuständen (vgl. S. 20) beträgt $1/(1-\hat{p}^{11}) = 6{,}2$ Wochen für Zustand 1 und $1/(1-\hat{p}^{22}) = 21{,}0$ Wochen für Zustand 2.

Aufschlußreich ist auch der Schätzwert für die (unbedingte) Wahrscheinlichkeit für Zustand 1 (vgl. S. 19):

$$\hat{\pi}_1 = \frac{1-\hat{p}^{22}}{(1-\hat{p}^{11})-(1-\hat{p}^{22})} = 0{,}228\,.$$

Dies entspricht näherungsweise der geschätzten Wahrscheinlichkeit dieses Zustandes im Mischungsmodell. Insgesamt ergibt sich also, daß sich alle Charakteristika des Mischungsmodells im komplexeren Markov-Modell wiederfinden. Zusätzlich erfaßt das Markov-Modell die Autokorrelation der quadrierten Werte, die sich im Markov-Modell in hohen geschätzten Verweilwahrscheinlichkeiten der vorwiegend durch die Volatilität charakterisierten Zustände widerspiegelt.

Dieser zusätzliche Erklärungsgehalt zeigt sich auch im Verlauf der geglätteten Wahrscheinlichkeiten für Zustand 1, der in Abbildung 10 dargestellt ist.

Da die Abfolge der Zustände in diesem Modell nicht mehr unabhängig ist, ist die Kurve glatter als die entsprechende des Mischungsmodells (vgl. Abbildung 9). Die einzelnen Phasen, in denen einer der beiden Zustände dominiert, treten hier deutlicher hervor. Da zur Berechnung der geglätteten Wahrscheinlichkeit für einen Zustand nicht nur der aktuelle Wert der Variablen verwendet wird, sondern *alle* Werte einfließen (vgl. Kapitel 4, Anhang A), können die Zustände hier besser unterschieden werden. Dies äußert sich z.B. darin, daß die geglätte-

[119] Zur Berechnung der ML-Schätzer vgl. den Anhang zu diesem Kapitel. Vgl. auch Fußnote 115.

ten Wahrscheinlichkeiten für das Markov-Modell (anders als beim Mischungsmodell) auch den Wert Null annehmen können, wie an Abbildung 10 ersichtlich ist.

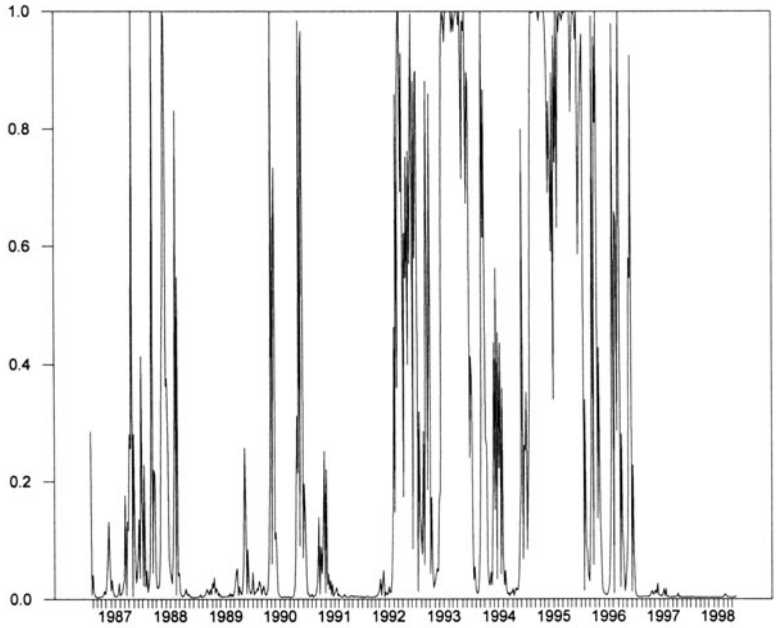

Abbildung 10: Geglättete Wahrscheinlichkeit für Zustand 1 (Markov-Modell)

Es ist zu vermuten, daß die Phasen mit hoher geglätteter Wahrscheinlichkeit *vor* der Erweiterung der Bandbreiten im August 1993 als Spekulation auf eine Abwertung des FF anzusehen sind, da der Wechselkurs in diesem Zeitraum mehrfach in die unmittelbare Nähe der Interventionsgrenzen gelangte. Interessant ist, daß diese Maßnahme die Unruhe auf den Devisenmärkten keineswegs sofort beendete. Diese Phase dauerte bis Ende 1996, mit einer kurzen, aber deutlichen Unterbrechung Ende 1994. Angesichts des großen Abstands des Wechselkurses von den neuen Interventionsgrenzen kann man aber nicht mehr davon ausgehen, daß in dieser Phase auf ein Realignment spekuliert wurde. Einen Ansatz zur Interpretation dieser lang andauernden Phase hoher Volatilität wird die Untersuchung der Monatsdaten des FF/DM-Kurses im Zusammenhang mit der PPP in Kapitel 10.2 ergeben.

Anhang: Verwendete Schätzverfahren

Alle im folgenden beschriebenen Berechnungen wurden mit dem Softwarepaket *RATS for Windows* (v. 4.30) der Firma *Estima* (Evanston, USA) durchgeführt, auf das sich auch die programmiertechnischen Hinweise beziehen.

1. Schwellenmodell

Hier liegt (wie bei den anderen beiden untersuchten Modellen auch) der Fall (3) aus Kapitel 4.1 (S. 37) vor. Wäre der Schwellenwert (und damit die Aufteilung der Beobachtungen auf die beiden Zustände) bekannt, so ließen sich die ML-Schätzer der Parameter, wie in Kapitel 4.2.1.2 erläutert, durch separate lineare Regressionen für die Beobachtungen, die jeweils einem Regime entsprechen, berechnen. Da der Schwellenwert unbekannt ist, wurde das in Kapitel 4.2.2 erläuterte Verfahren angewendet: Die eben erwähnten Regressionen wurden für je einen Schwellenwert zwischen je zwei benachbarten Beobachtungen von z (nämlich dem arithmetischen Mittel der beiden Werte) durchgeführt und daraus jeweils der Wert der Log-Likelihood nach der Formel auf S. 43 berechnet. Die angegebenen Schätzer sind diejenigen, für die dieser Wert maximal wurde. Der entsprechende Programmcode besteht also im wesentlichen aus einer Schleife über die Beobachtungen von z und zwei Regressionsanweisungen (nämlich dem LINREG-Befehl mit der Option SMPL zur Angabe der Beobachtungen, die den beiden Zuständen entsprechen), die bei jedem Durchlauf der Schleife ausgeführt werden.

2. Mischungsmodell

Diese Werte wurden gewonnen, indem sowohl der EM-Algorithmus als auch ein numerischer Algorithmus mit einem Gitterverfahren zur Bestimmung der Startwerte kombiniert wurde.

Der *EM-Algorithmus* läßt sich für Mischungsmodelle, wie in Kapitel 4.3.2 erläutert, relativ leicht umsetzen. Für den E-Schritt werden die geglätteten Wahrscheinlichkeiten der Zustandsvariablen benötigt, die mit Hilfe der auf S. 57 angegebenen Formel berechnet wurden. Für den M-Schritt wurden die in Kapitel 4.3.2.2.2 auf S. 59 definierten Matrizen (d.h. hier: Vektoren) $\mathbf{X}^i$ mit Hilfe der geglätteten Wahrscheinlichkeiten berechnet und damit die in diesem Kapitel unter **(3)** beschriebenen Regressionen durchgeführt. Der neue Parameter des Zustandsprozesses (d.h. hier p, die Wahrscheinlichkeit für Zustand 1) wurde mit Hilfe der auf S. 62 in Kapitel 4.3.2.2.4 angegeben Formel berechnet. Der E- und der M-Schritt wurden abwechselnd solange ausgeführt, bis keiner der Parameterwerte um mehr als 10^{-8} von dem entsprechenden im letzten Iterationsschritt abwich.

Das verwendete *numerische Verfahren* war der BHHH-(Berndt-Hall-Hall-Hausman-) Algorithmus in der Implementierung von *RATS for Windows* (v. 4.30).[120] Wie auf S. 51 in Kapitel 4.3 dargelegt wurde, hat die Likelihood für Mischungsmodelle eine einfache Produktgestalt (wobei hier p für $P(s_t = 1 \mid z_t)$ und $1-p$ für $P(s_t = 2 \mid z_t)$ einzusetzen ist). Deswegen lassen sich die ML-Schätzer in *RATS* sehr bequem mit dem Befehl MAXIMIZE (und der Option BHHH) berechnen. Es wurde dasselbe Konvergenzkriterium wie im Falle des EM-Algorithmus verwendet.

Als Startwerte für die Mittelwerte sowie für die Standardabweichungen wurden jeweils einmal die Schätzwerte des linearen Modells (Tabelle 1, S. 71) und die des Schwellenmodells (Tabelle 2, S. 74) in allen Kombinationen verwendet. Diese 4 Sätze von Startwerten wurden kombiniert mit den Startwerten 0,25, 0,5 und 0,75 für p. Daraus ergaben sich insgesamt 12 Kombinationen von Startwerten, die für das EM- und das numerische Verfahren zugrunde gelegt wurden. In 22 dieser 24 Rechnungen konvergierte der jeweilige Algorithmus bei den angegebenen Schätzwerten, bei den zwei übrigen Fällen an Werten, bei denen der Wert der Log-Likelihood niedriger war. Die angegebenen Werte repräsentieren deshalb mit großer Wahrscheinlichkeit das größte lokale Maximum.[121]

3. Markov-Modell

Für das Markov-Modell wurden analoge Schätzverfahren mit den gleichen Konvergenzkriterien verwendet.

Die geglätteten Wahrscheinlichkeiten und die bedingten gemeinsamen Wahrscheinlichkeiten aufeinanderfolgender Zustandsvariablen für den E-Schritt des *EM-Algorithmus* wurden mit Hilfe der im Anhang A von Kapitel 4 erläuterten Rekursion berechnet. Mit ihrer Hilfe wurden die neuen Parameterschätzer des M-Schrittes jeder Iteration nach den Formeln aus den Kapiteln 4.3.2.2.2 und 4.3.2.2.4 gewonnen. Dabei wurde für die Startwahrscheinlichkeiten angenommen, daß die Kette stationär ist.

Für die Implementierung des *numerischen Verfahrens* wurde von der auf S. 50 in Kapitel 4.3 angegebenen rekursiven Darstellung der Likelihood Gebrauch gemacht, die von GRAY [1996] (Appendix) vorgeschlagen wurde. Sie gestattet es, in *RATS* mit Hilfe des Befehls MAXIMIZE mit der Option RECURSIVE die ML-Schätzer für diesen Fall zu berechnen.

Die Startwerte wurden ähnlich wie für das Mischungsmodell bestimmt, wobei hier die Werte 0,25, 0,5 und 0,75 für p^{11} und p^{22} in beliebiger Kombination verwendet wurden, wo-

[120] Vgl. BERNDT u.a. [1974].

[121] Bei heteroskedastischen Modellen existiert kein globales Maximum, vgl. dazu Kapitel 6.1.2.

durch sich insgesamt 36 Sätze von Startwerten ergaben. In 65 der 72 Rechnungen mit den beiden Verfahren ergaben sich die angegebenen Parameterschätzer, in 5 Fällen Konvergenzpunkte mit kleinerem Wert der Log-Likelihood, und in 2 Fällen brach das Verfahren vorzeitig ab, weil eine Matrixinversion des BHHH-Verfahrens nicht durchgeführt werden konnte. Die angegebenen Werte repräsentieren deshalb ebenfalls mit großer Wahrscheinlichkeit das größte lokale Maximum.

6 Asymptotische Eigenschaften der ML-Schätzer und Tests

Aus der großen Zahl verschiedener Varianten von Regimewechselmodellen ergibt sich ein hohes Maß an Flexibilität in empirischen Anwendungen. Sie stellt den Anwender jedoch auch vor das Problem der Auswahl eines geeigneten Modells. Die Frage nach den in diesem Zusammenhang benötigten Tests ist eng verknüpft mit der Verteilung der ML-Schätzer. Diese läßt sich i.d.R. nicht exakt, sondern – wenn überhaupt – nur asymptotisch bestimmen. Auf diesen Themenbereich wird in Kapitel 6.1 eingegangen. Darauf aufbauend wird in Kapitel 6.2 die Testtheorie im Kontext von Regimewechselmodellen dargestellt.

6.1 Asymptotische Eigenschaften der ML-Schätzer

6.1.1 Schwellenmodelle

Für Schwellenmodelle ist die Situation grundsätzlich verschieden, je nachdem, ob der Schwellenwert bekannt ist oder nicht. Diese beiden Fälle werden deshalb im folgenden in getrennten Kapiteln behandelt.

6.1.1.1 Bekannter Schwellenwert

Es werden zuerst *homoskedastische* Modelle mit nichtstochastischen Regressoren ohne verzögert endogene Variablen betrachtet, also die Fälle **(2)** und **(4)** aus der Fallunterscheidung in Kapitel 4.1 (S. 37). Die Diskussion in Kapitel 4.2 hat gezeigt, daß sich diese Modelle als gewöhnliche lineare Modelle mit der Regressormatrix

$$\overline{\mathbf{X}}=\begin{pmatrix}\mathbf{X}^1 & \mathbf{X}^2 & \mathbf{W}\end{pmatrix}$$

und dem Parametervektor

$$\beta=\begin{pmatrix}\beta_1\\ \beta_2\\ \alpha\end{pmatrix}$$

auffassen lassen (in **(2)** gilt $\mathbf{W}=\mathbf{0}$ und $\alpha=\mathbf{0}$). Die ML-Schätzer für die Regressionsparameter sind die gewöhnlichen OLS-Schätzer für diese Modelle; die ML-Schätzer für die Störvarianz sind das $(T-2K-L)/T$-fache des gewöhnlichen Varianzschätzers (in **(2)** gilt $L=0$). Damit haben sie auch alle bekannten Eigenschaften dieser Schätzer.

Es gilt insbesondere:[122]

- $\hat{\beta}$ ist der *best linear unbiased estimator* (BLUE), $\hat{\beta} \sim N(\beta, \sigma^2(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1})$,
- $T\hat{\sigma}^2/\sigma^2 \sim \chi^2(T-2K-L)$, d.h. $E(\hat{\sigma}^2) = (1-\frac{2K+L}{T})\sigma^2$, $\hat{\sigma}^2$ ist also asymptotisch erwartungstreu,

  $Var(\hat{\sigma}^2) = 2\sigma^4 \frac{T-2K-L}{T^2}$, $\hat{\sigma}^2$ ist demnach MSE-*(Mean Square Error-)* konsistent[123] und damit insbesondere konsistent,
- $\hat{\beta}$ und $\hat{\sigma}^2$ sind unabhängig.

Ein erwartungstreuer Schätzer für die Varianz ist $\hat{s}^2 = T/(T-2K-L)\hat{\sigma}^2$.

Um Aussagen über die Eigenschaften der Schätzer für die Regressionsparameter in großen Stichproben machen zu können, muß man Annahmen über das Verhalten der Regressoren für große T treffen. Die übliche Annahme ist

$$\lim_{T\to\infty} \frac{1}{T}\overline{\mathbf{X}}'\overline{\mathbf{X}} = \mathbf{Q},$$

wobei $\mathbf{Q}$ eine positiv definite Matrix darstellt. Es gilt dann

- $\hat{\beta}$ ist MSE-konsistent.[124]

Die Annahme $\lim_{T\to\infty} 1/T(\overline{\mathbf{X}}'\overline{\mathbf{X}}) = \mathbf{Q}$ bedeutet in diesem Fall folgendes: Es ist

$$\overline{\mathbf{X}}'\overline{\mathbf{X}} = \begin{pmatrix} \mathbf{X}^{1\prime} \\ \mathbf{X}^{2\prime} \\ \mathbf{W}' \end{pmatrix} \begin{pmatrix} \mathbf{X}^1 & \mathbf{X}^2 & \mathbf{W} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X}^{1\prime}\mathbf{X}^1 & \mathbf{0} & \mathbf{X}^{1\prime}\mathbf{W}^1 \\ \mathbf{0} & \mathbf{X}^{2\prime}\mathbf{X}^2 & \mathbf{X}^{2\prime}\mathbf{W}^2 \\ (\mathbf{X}^{1\prime}\mathbf{W}^1)' & (\mathbf{X}^{2\prime}\mathbf{W}^2)' & \mathbf{W}'\mathbf{W} \end{pmatrix}.$$

In **(2)** mit $\mathbf{W} = \mathbf{0}$ wird also vorausgesetzt, daß die Quadrate (und gemischten Produkte) der Regressoren getrennt für $s_t = 1$ und $s_t = 2$ mit der Rate T, also dem Stichprobenumfang, wachsen.

---

[122] Vgl. z.B. GREENE [1997], S. 271.

[123] D.h. der Schätzer ist L^2-konvergent gegen den wahren Parameter.

[124] Die Annahme über die Momentenmatrix in großen Stichproben ist unnötig stark. Sie ist z.B. nicht erfüllt, falls einer der Regressoren ein linearer Zeittrend ist. In diesem Fall würde der Parameterschätzer noch schneller gegen den wahren Parameter konvergieren als in obigem Fall – er wäre *superkonsistent*, und zwar von der Ordnung in Wahrscheinlichkeit (*order in probability*) $T^{3/2}$. Man kann die Voraussetzungen, die erfüllt sein müssen, damit die Parameterschätzer konsistent sind, abschwächen. Einen hinreichenden Satz von Bedingungen bilden z.B. die sogenannten *Grenander-Bedingungen*, vgl. GREENE [1997], S. 274, oder JUDGE u.a. [1985], S. 162.

Für ein Regressionsmodell *ohne* Zustandswechsel lautet die entsprechende Bedingung $\lim_{T\to\infty}(1/T)\mathbf{X}'\mathbf{X} = \mathbf{Q}$. Es gilt

$$\begin{aligned}\mathbf{X}'\mathbf{X} &= \left(\mathbf{X}^{1\prime}+\mathbf{X}^{2\prime}\right)\left(\mathbf{X}^{1}+\mathbf{X}^{2}\right)\\ &= \mathbf{X}^{1\prime}\mathbf{X}^{1}+\mathbf{X}^{2\prime}\mathbf{X}^{2} \quad ,\end{aligned}$$

also ist die Voraussetzung $\lim_{T\to\infty}(1/T)\overline{\mathbf{X}}'\overline{\mathbf{X}} = \mathbf{Q}$ *stärker* als die entsprechende Bedingung für ein lineares Modell.

Falls diese quadratischen Terme zu verschiedenen Zeitpunkten in etwa den gleichen Wert haben, bedeutet die Annahme $\lim_{T\to\infty}(1/T)\overline{\mathbf{X}}'\overline{\mathbf{X}} = \mathbf{Q}$, daß der Anteil der Stichprobenwerte, der auf jeden der beiden Zustände entfällt, zeitlich ungefähr konstant sein muß. Das ist anschaulich leicht verständlich: Wenn ab einem gewissen Zeitpunkt für einen der beiden Zustände keine Werte mehr beobachtet werden, so kann man nicht erwarten, daß man die zu diesem Zustand gehörigen Regressionsparameter konsistent schätzen kann. In Fall **(4)** kommt als zusätzliche Bedingung hinzu, daß die quadratischen Terme von **W** mit der Rate T wachsen, und ebenso diejenigen von **X** und **W** (diese wiederum getrennt für $s_t = 1$ und $s_t = 2$).

Die allgemeine Theorie der ML-Schätzer macht genauere Aussagen über ihre asymptotische Verteilung. Sie besagt, daß – unter gewissen Regularitätsannahmen[125] – die ML-Schätzer asymptotisch erwartungstreu und normalverteilt sind, wobei die asymptotische Varianz-Kovarianz-Matrix die Inverse der Informationsmatrix, also die Cramér-Rao-Untergrenze für die Varianz-Kovarianz-Matrix eines jeden erwartungstreuen Schätzers der Parameter darstellt. ML-Schätzer sind demnach in solchen Fällen asymptotisch effizient.

Faßt man alle Parameter zu einem Vektor $\eta = (\beta_1, \beta_2, \alpha, \sigma^2)$ zusammen, so konvergiert das $T^{1/2}$-fache dieses Vektors also *in Verteilung* gegen die angesprochene Normalverteilung:

$$\sqrt{T}(\hat{\eta}_n - \eta_0) \xrightarrow{d} N(0, I(\eta_0)^{-1}) \ ,$$

mit

$$I(\eta_0) = -E_{\eta_0}\left(\frac{\partial^2 \log f(\mathbf{y};\eta_0)}{\partial\eta\partial\eta'}\right).$$

Die Informationsmatrix lautet in dem hier betrachteten Fall

[125] Diese Annahmen sind – ähnlich wie die des zentralen Grenzwertsatzes – kompliziert und i.a. schwer nachzuprüfen, vgl. z.B. GOURIEROUX und MONFORT [1995a], S. 183 ff. Die Autoren weisen darauf hin, daß die Aussage für Regressionsmodelle nicht richtig sein muß (vgl. S. 193). Sie zeigen jedoch, daß sie gilt, wenn die obige Annahme über das Verhalten der Momentenmatrix für wachsende Stichproben erfüllt ist.

$$I(\eta_0) = \begin{pmatrix} \frac{1}{\sigma^2}\mathbf{X}^{1\prime}\mathbf{X}^1 & \mathbf{0}^{a)} & \frac{1}{\sigma^2}\mathbf{X}^{1\prime}\mathbf{W}^1 & \mathbf{0}^{b)} \\ \mathbf{0}^{a)} & \frac{1}{\sigma^2}\mathbf{X}^{2\prime}\mathbf{X}^2 & \frac{1}{\sigma^2}\mathbf{X}^{2\prime}\mathbf{W}^2 & \mathbf{0}^{b)} \\ \frac{1}{\sigma^2}\left(\mathbf{X}^{1\prime}\mathbf{W}^1\right)' & \frac{1}{\sigma^2}\left(\mathbf{X}^{2\prime}\mathbf{W}^2\right)' & \frac{1}{\sigma^2}\mathbf{W}'\mathbf{W} & \mathbf{0}^{b)} \\ \mathbf{0}^{b)} & \mathbf{0}^{b)} & \mathbf{0}^{b)} & \frac{T}{2\sigma^4} \end{pmatrix},$$

wobei an den mit $^{a)}$ markierten Stellen eine **0** steht, weil

$$\frac{\partial^2 \log f(\mathbf{y};\eta_0)}{\partial\beta_1\partial\beta_2'} = \frac{\partial^2 \log f(\mathbf{y};\eta_0)}{\partial\beta_2\partial\beta_1'} = \mathbf{0}\ ,$$

und an den mit $^{b)}$ markierten Stellen, weil (zum Beispiel)

$$\begin{aligned} E_{\eta_0}\left(\frac{\partial^2 \log f(\mathbf{y};\eta_0)}{\partial\beta_1\partial\sigma^2}\right) &= E_{\eta_0}\left(-\frac{1}{\sigma^4}\mathbf{X}^{1\prime}\left(\mathbf{y}^1 - \mathbf{X}^1\beta_1 - \mathbf{W}^1\alpha\right)\right) \\ &= E_{\eta_0}\left(-\frac{1}{\sigma^4}\mathbf{X}^{1\prime}\mathbf{u}^1\right) \\ &= -\frac{1}{\sigma^4}\mathbf{X}^{1\prime}E_{\eta_0}\mathbf{u}^1 \\ &= \mathbf{0}. \end{aligned}$$

Die asymptotische Varianz-Kovarianz-Matrix lautet

$$I(\eta_0)^{-1} = \begin{pmatrix} \sigma^2(\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\sigma^4}{T} \end{pmatrix}.$$

Diese asymptotische Verteilung läßt sich für den homoskedastischen Fall auch ohne die allgemeine Theorie von ML-Schätzern herleiten: Die Schätzer für die Regressionsparameter sind, wie oben dargelegt, schon in endlichen Stichproben normalverteilt mit genau dieser Varianz-Kovarianz-Matrix (d.h. der ersten $(2K+L)$-Untermatrix der Informationsmatrix), sie sind schon in endlichen Stichproben unabhängig von dem Schätzer für die Varianz, und da dieser in endlichen Stichproben χ^2-verteilt ist, folgt die asymptotische Normalverteilung mit der angegebenen Varianz hier auch aus dem Zentralen Grenzwertsatz.

Im *heteroskedastischen* Fall kann man in der Regel nur Aussagen über die asymptotischen Eigenschaften der ML-Schätzer machen. Eine Ausnahme hiervon stellt der Fall **(3)** dar, bei dem alle Regressionsparameter zustandsabhängig sind. Wie die Diskussion in Kapitel 4.2 gezeigt hat, gilt hier wie im entsprechenden homoskedastischen Fall

$$\begin{aligned}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} = \hat{\beta} &= (\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\mathbf{y}\\ &= (\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'(\overline{\mathbf{X}}\beta + \mathbf{u})\\ &= \beta + (\overline{\mathbf{X}}'\overline{\mathbf{X}})^{-1}\overline{\mathbf{X}}'\mathbf{u}.\end{aligned}$$

Es gilt

- $\hat{\beta}$ ist der *best linear unbiased estimator* (BLUE),

$$\hat{\beta} \sim N(\beta, \begin{pmatrix}\sigma_1^2(\mathbf{X}^{1\prime}\mathbf{X}^1)^{-1} & \mathbf{0}\\ \mathbf{0} & \sigma_2^2(\mathbf{X}^{2\prime}\mathbf{X}^2)^{-1}\end{pmatrix},$$

- $T_i\hat{\sigma}_i^2/\sigma_i^2 \sim \chi^2(T_i - K)$, d.h. $\hat{\sigma}_i^2$ ist asymptotisch erwartungstreu und MSE-konsistent,
- $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$ sind unabhängig.

Erwartungstreue Schätzer für die Varianzen stellen hier $\hat{s}_i^2 = T/(T-K)\hat{\sigma}_i^2$ dar.

Für diesen und die übrigen heteroskedastischen Fälle **(1)** und **(5)** gelten unter gewissen Voraussetzungen[126] wieder die analogen Aussagen über die asymptotischen Eigenschaften der ML-Schätzer. Die wichtigste Annahme ist hier

$$\lim_{T\to\infty}\frac{1}{T}\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}} = \mathbf{Q},$$

wobei $\mathbf{Q}$ wieder eine positiv definite Matrix darstellt. Es gilt

$$\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}} = \begin{pmatrix}\frac{1}{\sigma_1^2}\mathbf{X}^{1\prime}\mathbf{X}^1 & \mathbf{0} & \frac{1}{\sigma_1^2}\mathbf{X}^{1\prime}\mathbf{W}^1\\ \mathbf{0} & \frac{1}{\sigma_2^2}\mathbf{X}^{2\prime}\mathbf{X}^2 & \frac{1}{\sigma_2^2}\mathbf{X}^{2\prime}\mathbf{W}^2\\ \left(\frac{1}{\sigma_1^2}\mathbf{X}^{1\prime}\mathbf{W}^1\right)' & \left(\frac{1}{\sigma_2^2}\mathbf{X}^{2\prime}\mathbf{W}^2\right)' & \frac{1}{\sigma_1^2}\mathbf{W}^{1\prime}\mathbf{W}^1 + \frac{1}{\sigma_2^2}\mathbf{W}^{2\prime}\mathbf{W}^2\end{pmatrix}.$$

Der wichtigste Unterschied zwischen der Annahme $\lim_{T\to\infty}(1/T)\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}} = \mathbf{Q}$ und der entsprechenden im homoskedastischen Fall betrifft die rechte untere Blockmatrix: Die mit den Kehrwerten der zugehörigen Varianzen gewichteten (gemischten) Quadrate von Elementen von $\mathbf{W}$ aus den beiden Zuständen müssen mit der Rate T wachsen. Hinreichend (aber nicht notwendig) hierfür ist z.B., daß

$$\lim_{T\to\infty}\frac{1}{T}\mathbf{W}^{1\prime}\mathbf{W}^1 \text{ und } \lim_{T\to\infty}\frac{1}{T}\mathbf{W}^{2\prime}\mathbf{W}^2$$

getrennt existieren und positiv definit sind.

[126] Vgl. MAGNUS [1978], S. 294 f.

Faßt man die Parameter jetzt zu dem Vektor $\eta = (\beta_1, \beta_2, \alpha, \sigma_1^2, \sigma_2^2)$ zusammen, so gilt unter dieser Annahme wieder

$$\sqrt{T}(\hat{\eta}_n - \eta_0) \xrightarrow{d} N(0, I(\eta_0)^{-1}) \ .$$

Die asymptotische Varianz-Kovarianz-Matrix lautet in diesem Fall

$$I(\eta_0)^{-1} = \begin{pmatrix} \left(\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dfrac{2\sigma_1^4}{T_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dfrac{2\sigma_2^4}{T_2} \end{pmatrix} .$$

Die ML-Schätzer sind also auch hier konsistent, sowie asymptotisch erwartungstreu, normalverteilt und effizient. Asymptotisch sind die Schätzer für die Regressionsparameter und die für die Varianzen (wie im gewöhnlichen linearen Modell) unabhängig. Die asymptotischen Varianzen von $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$ sind die Werte, die man erhält, wenn man die Ergebnisse für das gewöhnliche lineare Modell jeweils auf die beiden Stichproben mit $s_t = 1$ und $s_t = 2$ überträgt.

Sind die Regressoren stochastisch, so sind die meisten dieser Aussagen weiterhin gültig, wenn man die stochastischen Annahmen etwas modifiziert.[127] Nimmt man an, daß **u** und **X** stochastisch unabhängig sind und $E(\mathbf{u}|\mathbf{X}) = 0$ gilt, so sind $\hat{\beta}$ und $\hat{s}^2$ bzw. $\hat{s}_i^2$ $(i = 1,2)$ zumindest in den Fällen **(1)**, **(2)** und **(4)** erwartungstreu. Die Varianz von $\hat{\beta}$ ist in den homoskedastischen Fällen $\sigma^2 E\left[(\overline{\mathbf{X}}'\overline{\mathbf{X}}^{-1})\right]$, die Verteilung von $\hat{\sigma}^2$ bzw. $\hat{\sigma}_i^2$ $(i = 1,2)$ ist identisch mit der im Fall nichtstochastischer Regressoren, und $\hat{\beta}$ und $\hat{\sigma}^2$ bzw. $\hat{\sigma}_i^2$ $(i = 1,2)$ sind unkorreliert (aber nicht notwendig unabhängig).

Alle asymptotischen Eigenschaften, also die Konsistenz sowie die asymptotische Erwartungstreue, Normalverteilung und Effizienz, gelten unverändert, wenn man die Annahme über die Momentenmatrix in großen Stichproben ersetzt durch

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\overline{\mathbf{X}}'\overline{\mathbf{X}} = \mathbf{Q} \ ,$$

wobei **Q** wieder eine (konstante) positiv definite Matrix darstellt.[128]

Bedingte ML-Schätzer haben unter bestimmten Voraussetzungen – die aber i.a. nicht leicht zu verifizieren sind – dieselben asymptotischen Eigenschaften wie ML-Schätzer.[129] Es ist

[127] Vgl. GREENE [1997], S. 321 ff.

[128] Vgl. z.B. GREENE [1997], S. 322 ff., oder JUDGE u.a. [1985], S. 164 ff.

deshalb zu vermuten, daß die asymptotischen Aussagen auch für die übrigen heteroskedastischen Fälle **(3)** und **(5)** gültig sind, wenn man hier voraussetzt, daß

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}}=\mathbf{Q}\ ,$$

wobei **Q** wieder eine (konstante) positiv definite Matrix darstellt.

Für Zeitreihenmodelle schließlich bleiben die Konsistenz, die asymptotische Effizienz und die asymptotische Normalverteilung erhalten, falls der Prozeß stationär ist.[130] Die Informationsmatrix hat eine ähnliche Gestalt wie im Falle nichtstochastischer Regressoren, wobei hier die Momentenmatrix durch die Varianz-Kovarianz-Matrix von $y_1,\ldots,y_p$ zu ersetzen ist.[131]

6.1.1.2 Unbekannter Schwellenwert

Wenn die Aufteilung der Stichprobe auf die beiden Zustände geschätzt werden muß, lassen sich i.d.R. keine Aussagen über die exakte Verteilung der ML-Schätzer mehr machen. Auch die asymptotische Theorie der ML-Schätzer ist hier nicht ohne weiteres anwendbar, da infolge der Unstetigkeiten in der Likelihood-Funktion die Regularitätseigenschaften verletzt sind, die für diese asymptotischen Eigenschaften Voraussetzung sind. Dennoch läßt sich zeigen, daß die ML-Schätzer auch in diesem Fall gute Eigenschaften haben. In CHAN [1988] wird gezeigt, daß für reine Zeitreihen-Schwellenmodelle (SETAR-Modelle) mit unbekanntem Schwellenwert i.a.

- die ML-Schätzer stark konsistent,[132] also auch (schwach) konsistent sind,
- der ML-Schätzer für c sogar superkonsistent von der Ordnung in Wahrscheinlichkeit T ist, d.h.

$$\operatorname*{plim}_{T\to\infty}T(\hat{c}-c)=0\ ,$$

- die ML-Schätzer für die Regressionsparameter asymptotisch normalverteilt sind, mit derselben Varianz-Kovarianz-Matrix wie im Falle eines bekannten Schwellenwertes,
- die ML-Schätzer für die Regressionsparameter und die für den Schwellenwert asymptotisch unabhängig sind.[133]

---

[129] Vgl. z.B. GOURIEROUX und MONFORT [1995a], S. 198. ff, und GOURIEROUX und MONFORT [1995b], Kapitel 24.

[130] Vgl. JUDGE u.a. [1985], S. 238 f.

[131] Vgl. HARVEY [1993], S. 130 f.

[132] D.h. die Schätzer konvergieren mit wachsendem Stichprobenumfang *fast sicher* gegen die wahren Parameter.

[133] Vgl. hierzu auch TONG [1990], Kapitel 5.5.3.

Es ist zu vermuten, daß diese Aussagen unter geeigneten Annahmen über die Regressoren ($\lim_{T\to\infty}(1/T)\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}} = \mathbf{Q}$ bzw. $\operatorname{plim}_{T\to\infty}(1/T)\overline{\mathbf{X}}'\Omega^{-1}\overline{\mathbf{X}} = \mathbf{Q}$) auch auf die hier betrachteten Fälle mit exogenen Regressoren ausgedehnt werden können. Der ML-Schätzer für den Schwellenwert hingegen ist nicht asymptotisch normalverteilt.[134]

6.1.2 Mischungs- und Markov-Modelle

Über die exakte Verteilung der ML-Schätzer kann man für Mischungs- und Markov-Modelle i.a. keine Aussagen machen. Auch die allgemeine Theorie der ML-Schätzer läßt sich hier nicht ohne weiteres anwenden, da eine wichtige Voraussetzung, die man benötigt (falls sie nicht durch andere Annahmen ersetzt werden kann, wie im Falle eines gewöhnlichen Regressionsmodells oder eines Modells mit beobachteten Zuständen[135]), hier nicht erfüllt ist, nämlich die identische Verteilung der Stichproben zu verschiedenen Zeitpunkten.

Ein besonderes Problem tritt im heteroskedastischen Fall auf: Die Likelihood-Funktion ist unbeschränkt. Sie lautet z.B. für Markov-Modelle (vgl. Kapitel 4.3 , S. 49)

$$f(\mathbf{y} \mid \mathbf{X}, \mathbf{W}, Y_0) = \sum_{\mathbf{s}} f(\mathbf{y}, \mathbf{s} \mid \mathbf{X}, \mathbf{W}, Y_0)$$

$$= \sum_{\mathbf{s}} \left[\left(\prod_{t=2}^{T} f(y_t \mid s_t, \mathbf{X}_t, \mathbf{W}_t) P(s_t \mid s_{t-1}) \right) f(y_1 \mid s_1, \mathbf{X}_1, \mathbf{W}_1) P(s_1 \mid Y_0, \mathbf{X}, \mathbf{W}) \right].$$

Wählt man die Regressionsparameter für einen der beiden Zustände, etwa α und β_1, so, daß das Residuum zu einem beliebigen (festen) Zeitpunkt t_0 gleich 0 ist, so gilt

$$\lim_{\sigma_1 \to 0} f(y_{t_0} \mid s_{t_0} = 1, \mathbf{X}_{t_0}, \mathbf{W}_{t_0}) = \lim_{\sigma_1 \to 0} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_{t_0} - \mathbf{W}_{t_0}\alpha - \mathbf{X}_{t_0}\beta_1)^2}{2\sigma_1^2}} = \lim_{\sigma_1 \to 0} \frac{1}{\sqrt{2\pi}\sigma_1} = \infty \quad .$$

Wählt man nun $\sigma_2 \neq 0$, so ist

$$f(y_t \mid s_t = 2, \mathbf{X}_t, \mathbf{W}_t) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_t - \mathbf{W}_t\alpha - \mathbf{X}_t\beta_1)^2}{2\sigma_1^2}} > 0 \quad ,$$

und für $\mathbf{s} = (s_t)_{t=1,\ldots,T}$ mit $s_t = \begin{cases} 1 \text{ für } t = t_0 \\ 2 \text{ für } t \neq t_0 \end{cases}$, $P(s_t \mid s_{t-1}) \neq 0$ $(t = 2,\ldots,T)$ und $P(s_1 \mid Y_0, \mathbf{X}, \mathbf{W}) \neq 0$ gilt

$$\lim_{\sigma_1 \to 0} f(\mathbf{y}, \mathbf{s} \mid \mathbf{X}, \mathbf{W}, Y_0) = \infty \quad .$$

[134] Zur seiner asymptotischen Verteilung vgl. CHAN [1988] und TONG [1990], S. 293 ff.

[135] Vgl. Kapitel 6.1.

Damit gilt auch

$$\lim_{\sigma_1 \to \infty} f(\mathbf{y} \mid \mathbf{X}, \mathbf{W}, Y_0) = \lim_{\sigma_1 \to \infty} \sum_{\mathbf{s}} f(\mathbf{y}, \mathbf{s} \mid \mathbf{X}, \mathbf{W}, Y_0) = \infty \ .$$

Die größte (nämlich unendlich große) Likelihood hat folglich ein Modell ohne Residuum für genau eine Beobachtung, das aber sicher nicht sonderlich geeignet ist, reale Phänomene zu beschreiben. In empirischen Anwendungen hat sich jedoch gezeigt, daß man sinnvolle Schätzergebnisse erhält, wenn man als Schätzwert ein lokales Maximum der Likelihood-Funktion (möglichst das mit dem größten Wert der Likelihood unter allen lokalen Maxima) wählt. Tatsächlich läßt sich für *homogene Modelle* ohne verzögert endogene Regressoren zeigen, daß[136]

- es für große T nur ein einziges lokales Maximum der Likelihood-Funktion gibt, und daß der zugehörige Schätzer konsistent ist
- dieser Schätzer asymptotisch erwartungstreu und normalverteilt ist, wobei die asymptotische Varianz-Kovarianz-Matrix gleich der Inversen der Informationsmatrix ist.

Für inhomogene Modelle mit verzögert endogenen Regressoren existieren keine derart geschlossenen Ergebnisse. Dennoch gibt es Resultate, die die Annahme gerechtfertigt erscheinen lassen, daß die ML-Schätzer auch in diesen Fällen ähnlich gute Eigenschaften aufweisen.[137] Deswegen gehen die meisten Autoren von der Gültigkeit dieser Eigenschafen aus, auch wenn sie (bisher) nicht in allen Fällen nachgewiesen werden konnten.

6.2 Tests

Bei der Vielzahl von Regimewechselmodellen, die in den vorangegangenen Kapiteln behandelt wurden, ist es in Anwendungen unverzichtbar, über ein objektives Entscheidungskriterium in Form von statistischen Tests zu verfügen, die Auskunft darüber geben, *ob* ein Regimewechselmodell die vorliegenden Daten signifikant besser beschreibt als ein gewöhnliches (lineares) Modell, und wenn ja, *welches* Regimewechselmodell angebracht ist.

Die erste Frage führt auf Tests von linearen gegen Regimewechselmodelle. Diese Tests, die am Anfang jeder ökonometrischen Untersuchung mit Hilfe von Regimewechselmodellen stehen sollten, werden (nach einer kurzen Einführung in die Problematik asymptotischer Tests im Zusammenhang von Regimewechselmodellen in Kapitel 6.2.1) in Kapitel 6.2.2 behandelt.

Die zweite Frage läßt sich noch einmal unterteilen in zwei verschiedene Problemstellungen. Zum einen kann man daran interessiert sein, zwischen Regimewechselmodellen mit verschie-

[136] Für Mischungsmodelle vgl. KIEFER [1978], für Markov-Modelle vgl. LEROUX [1992].

[137] Vgl. LEE [1991], chapter 5.

denen Mechanismen, die die Zustandswechsel bestimmen, auszuwählen. Typische Fragen in diesem Kontext sind etwa, ob die Zustände Permanenz zeigen, die auf ein Markov-Modell hinweist, oder ob ein Mischungsmodell mit unabhängigen Zuständen die Daten (fast) genau so gut beschreibt; oder, ob der Einfluß eines beobachteten Zustandsindikators auf die Abfolge der Zustände signifikant ist, so daß ein inhomogenes (Mischungs- oder Markov-)Modell deutlich angebrachter erscheint als das entsprechende homogene Modell. Diese Testsituationen werden in Kapitel 6.2.3 untersucht.

Zum anderen kann man Hypothesen über die Regressionsparameter eines Regimewechselmodells testen, d.h. zwischen Modellen auswählen, für die die Dynamik des Zustandsprozesses dieselbe ist, die sich aber in der Spezifikation in den einzelnen Zuständen unterscheiden, also in der Auswahl der Regressoren, der autoregressiven Dynamik oder der Zustandsabhängigkeit einzelner Regressionsparameter bzw. der Störvarianz. Geeignete Tests dieser Hypothesen werden in Kapitel 6.2.4 behandelt.[138]

In Kapitel 6.2.5 schließlich wird ein Verfahren beschrieben, mit dem man zwischen *getrennten* Modellen, von denen sich also keines durch Restriktion aus dem anderen ergibt, testen kann.

6.2.1 Asymptotische Tests im Kontext von Regimewechselmodellen

In vielen der angesprochenen Testsituationen werden die üblichen asymptotischen Tests, d.h. der Likelihood-Quotienten-(LQ-)Test, der Wald- und der Lagrange-Multiplikatoren-(LM-)Test verwendet. Diese Tests besitzen die bekannten asymptotischen Optimalitätseigenschaften: Sie sind i.a. asymptotisch lokal *most powerful*, und sie sind konsistent (d.h. ihre Macht oder *power* konvergiert für wachsende Stichproben gegen 1), wenn die ML-Schätzer konsistent sind, was ja, wie die Diskussion in Kapitel 6.1 gezeigt hat, i.d.R. der Fall ist.[139] In den meisten Situationen, in denen diese Tests verwendet werden, ist zwar nur wenig über die Güte der Approximation der Verteilung der Teststatistiken durch die entsprechende χ^2-Verteilung in Abhängigkeit vom Stichprobenumfang bekannt (also über das tatsächliche

[138] In der Literatur sind darüber hinaus noch *diagnostische Tests* vorgeschlagen worden, also allgemeine Anpassungstests, mit denen nach Wahl eines Regimewechselmodells überprüft werden kann, ob neben dem Zustandswechsel noch weitere Effekte vorliegen, die einer besonderen Modellierung bedürfen, wie etwa Autokorrelation der Residuen oder verallgemeinerte ARCH-Effekte. Vgl. hierzu HAMILTON [1996].

[139] Vgl. hierzu z.B. ENGLE [1984], S. 796 ff., oder GOURIEROUX und MONFORT [1995b], Kapitel 17.2.1-17.2.3.

Niveau der Tests).[140] Dennoch werden sie häufig angewendet, da sie in vielen Fällen die einzigen allgemein akzeptierten Testverfahren darstellen.

Im allgemeinen sind diese Tests unter den gleichen Bedingungen, unter denen die ML-Schätzer asymptotisch normalverteilt sind, gültig (d.h. die Teststatistiken sind asymptotisch χ^2-verteilt). Wie in Kapitel 6.1 erläutert wurde, liegt eine asymptotische Normalverteilung in manchen Fällen (z.B. unter bestimmten Voraussetzungen bei bekannter Regimeklassifikation) vor. In anderen Fällen (z.B. bei abhängigen Beobachtungen wie im Fall von Markov-Modellen) liegen (noch) keine strengen Beweise dafür vor. Es wird aber in der Literatur von der Gültigkeit der Normalverteilung ausgegangen, und die Verwendung der asymptotischen Tests ist weit verbreitet. In wieder anderen Fällen sind die ML-Schätzer einiger Parameter definitiv *nicht* asymptotisch normalverteilt (wie etwa im Falle eines Schwellenmodells mit unbekanntem Schwellenwert der ML-Schätzer dieses Schwellenwertes). Hier können die asymptotischen Tests nicht in ihrer üblichen Form angewendet werden.

Eine weitere Komplikation ergibt sich bei den Linearitätstests dadurch, daß in den meisten Fällen einer oder mehrere Modellparameter der Alternative unter der Nullhypothese *nicht identifiziert* sind. Dies führt dazu, daß die üblichen Teststatistiken nicht die erwartete asymptotische χ^2-Verteilung besitzen.[141] Auf diese Schwierigkeiten wird in den folgenden Kapiteln genauer eingegangen, und es werden möglichst praktikable Modifikationen der gängigen asymptotischen Tests vorgeschlagen.

Unter der Voraussetzung, daß die ML-Schätzer asymptotisch normalverteilt sind, sind die drei Tests asymptotisch äquivalent. Für endliche Stichproben kann und wird es aber vorkommen, daß sie zu unterschiedlichen Ergebnissen führen, so daß die Auswahl des zu verwendenden Testverfahrens von Bedeutung ist. Die Testtheorie liefert keine grundsätzlichen Anhaltspunkte für eine Entscheidung für einen dieser drei Tests, weswegen man sie aus praktischen Erwägungen heraus treffen wird. Ein Vorteil des LM-Tests bzw. des Wald-Tests ist, daß man zur Berechnung der Teststatistik die Likelihood-Funktion nur unter der Nullhypothese bzw. unter der Alternative zu maximieren braucht, während der LQ-Test die Berechnung *beider* Maxima erfordert. Andererseits ist zur Berechnung der LM- bzw. der Wald-Statistik die Bestimmung eines Schätzers der Informationsmatrix an dem jeweiligen Maximum nötig. Dies kann ein komplexes Problem darstellen. Wenn man jedoch einen Schätzalgorithmus verwen-

[140] In einigen wenigen Fällen läßt sich die Verteilung der Teststatistiken (oder dazu äquivalenter, d.h. streng monoton von ihnen abhängenden Größen) explizit angeben. Darauf wird im folgenden an den entsprechenden Stellen genauer eingegangen.

[141] Vgl. zu diesem Problem DAVIES [1977, 1987], ENGLE [1984], S. 823 f., und ANDREWS und PLOBERGER [1994].

det, der mit solchen geschätzten Informationsmatrizen arbeitet (vgl. 4.3.1), erhält man nach Konvergenz des Verfahrens automatisch einen solchen Schätzer der Informationsmatrix an der Stelle des Maximums, und kann damit die Teststatistiken bequem berechnen. Allerdings ergibt sich aus der Möglichkeit, zwischen den verschiedenen Schätzern für die Informationsmatrix auszuwählen, die i.a. auch zu verschiedenen Werten der Teststatistiken führen, eine zusätzliche Mehrdeutigkeit in dem jeweiligen Testproblem.[142]

In der Literatur über Regimewechselmodelle ist eine leichte Bevorzugung des LQ-Tests gegenüber den anderen beiden asymptotischen Tests zu beobachten. Dies zeigt sich besonders daran, daß in Fällen, in denen die asymptotische Verteilung der Teststatistiken fraglich ist, in aller Regel Simulationen für die LQ-Teststatistik durchgeführt werden. Das mag an der oben geschilderten Notwendigkeit der Bestimmung von geschätzten Informationsmatrizen für die anderen beiden Tests liegen, oder daran, daß der LQ-Test (in Verbindung mit dem Neymann-Pearson-Lemma) den ursprünglichsten und intuitiv einleuchtendsten Zugang zum Testproblem darstellt, während die anderen beiden Tests gewöhnlich als Approximationen des LQ-Tests angesehen werden. Ein weiterer Grund ist, daß der LQ-Test im Zusammenhang mit dem Problem nicht identifizierter Parameter relativ robust ist, während die Verteilung der Wald- und der LM-Teststatistik offensichtlich extreme Parameterabhängigkeit aufweist.[143]

Aus diesen Gründen wird im folgenden an einigen Stellen der LQ-Test verwendet, an denen auch die anderen beiden Tests hätten in Betracht gezogen werden können. Dadurch soll eine größtmögliche Vergleichbarkeit der hier präsentierten Ergebnisse mit denen aus der Literatur erreicht werden.

6.2.2 Tests linearer Modelle gegen Regimewechselmodelle

Alle Regimewechselmodelle enthalten als Spezialfälle lineare Modelle *ohne* Zustandswechsel. Man erhält sie z.B. durch die Restriktion, daß alle Regressionsparameter in den verschiedenen Zuständen identisch sind. Es hat sich die Auffassung durchgesetzt, daß auf ein Modell mit mehreren Zuständen nur geschlossen werden sollte, wenn die Abweichung von dem entsprechenden gewöhnlichen Modell mit einem Zustand signifikant ist. Damit stellt sich die Frage nach Tests der Nullhypothese eines gewöhnlichen Regressions- oder Zeitreihenmodells gegen die Alternative eines Modells mit mehreren Zuständen.

---

[142] Gelegentlich wird auch, basierend auf der asymptotischen Normalverteilung der Parameter, mit asymptotischen *t*-Tests von Hypothesen über *einzelne* Parameter mit Hilfe von geschätzten asymptotischen Varianz-Kovarianz-Matrizen argumentiert. Vgl. hierzu den Anhang von Kapitel 7.

[143] Vgl. hierzu LEE und CHESHER [1986] und DUFOUR [1995].

Den allgemeinen Linearitätstests[144] sind wegen ihrer vergleichsweise geringen Macht grundsätzlich solche mit klar spezifizierter Alternative vorzuziehen. Hier bieten sich die üblichen asymptotischen Tests an. Diese können ohne Modifikation nur im Fall von Schwellenmodellen mit bekanntem Schwellenwert angewendet werden, die deswegen in Kapitel 6.2.2.1.1 gesondert behandelt werden.

In den übrigen Fällen tritt das bereits erwähnte Problem auf, daß einige der Modellparameter der Alternative unter der Nullhypothese nicht identifiziert sind. Diese sogenannten *Störparameter* (*nuisance parameters*) sind bei Schwellenmodellen der Schwellenparameter, bei Mischungsmodellen die Aufenthaltswahrscheinlichkeit und bei Markov-Modellen die Start- und Übergangswahrscheinlichkeiten bzw. (im inhomogenen Fall) die Parameter, die die Abhängigkeit dieser Größen von dem Zustandsindikator festlegen. Damit sind die Regularitätsvoraussetzungen nicht erfüllt, die sicherstellen, daß die Teststatistiken der asymptotischen Tests asymptotisch $\chi^2(n)$-verteilt sind (wobei n die Anzahl der linear unabhängigen Restriktionen bezeichnet). Deswegen können diese Tests nicht unverändert verwendet werden. In den Kapiteln 6.2.2.1.2 und 6.2.2.2 werden verschiedene Modifikationen dieser Tests für Schwellenmodelle mit unbekanntem Schwellenwert und für Mischungs- und Markov-Modelle diskutiert.

Es gibt einige theoretische Resultate über die Verteilung der LQ-Statistik in diesen Fällen. Nachdem die verschiedenen Modelltypen längere Zeit getrennt voneinander untersucht wurden, ist in der letzten Zeit der Versuch unternommen worden, einen einheitlichen Zugang zu dem Problem der Linearitätstests für Markov-Modelle und Schwellenmodelle mit unbekanntem Schwellenwert zu entwickeln. In GARCIA [1998] wird für einfach strukturierte Modelle gezeigt, daß in beiden Fällen die LQ-Statistik gegen ein Funktional einer Brownschen Brücke konvergiert. Dies gilt aber nur, falls man geeignete a-priori-Restriktionen für den Wert des Störparameters einführt. Man muß etwa p^{11} und p^{22} auf ein kompaktes Intervall $[\alpha,1-\alpha]$ beschränken, das echt in $[0,1]$ enthalten ist, bzw. den Schwellenwert c auf den Bereich zwischen dem α- und dem $(1-\alpha)$-Quartil der empirischen Verteilung von z_t, wobei $\alpha = 0{,}15$ üblich ist.[145]

Da diese Grenzverteilung aber abhängig von den wahren Parametern ist, ist der Anwender nach wie vor auf Simulationen der Verteilung der Teststatistik angewiesen. Im folgenden wird

[144] Zu diesen auch als *Portmanteau*-Tests bezeichneten Tests vgl. TONG [1990], S. 221-229 und TERÄSVIRTA, TJØSTHEIM und GRANGER [1994], S. 2930-2933.

[145] Der Beweis hierzu wird bei CARRASCO [1997] auf komplexere Modelle ausgedehnt. Die Arbeit von GARCIA [1998] kursierte bereits seit 1995 als *Cahier de Recherche* Nr. 1095 der Université de Montréal, sie ist also tatsächlich älter als die von Carrasco.

daher vorwiegend auf Näherungsformeln, eigene Simulationen für LQ-Tests und Simulationsergebnisse aus der Literatur eingegangen. Es wird gezeigt, daß sie sich in einfachen Faustregeln zusammenfassen lassen, die für empirische Anwendungen nützlich sind.

6.2.2.1 Schwellenmodelle

6.2.2.1.1 Bekannter Schwellenwert

In Kapitel 4.2.1 wurde dargelegt, daß im Falle eines bekannten Schwellenwertes, also bei beobachtbaren Zuständen, Schwellenmodelle als gewöhnliche Regressionsmodelle mit spezieller Regressormatrix angesehen werden können.[146]

Die Nullhypothese eines Modells ohne Zustandswechsel entspricht dabei der Restriktion, daß die regimeabhängigen Parameter (inklusive der Störvarianzen) in den beiden Zuständen gleich sind. Dies ist eine lineare Hypothese über die Regressionsparameter, die im Falle eines homoskedastischen Modells mit exogenen Regressoren mit dem üblichen F-Test getestet werden kann, der in diesem Fall dem *exakten* LQ-Test (für endliche Stichproben) entspricht.

Ist die Alternative ein heteroskedastisches Modell, so kann man entweder zuerst den eben erwähnten F-Test durchführen und danach, falls die Nullhypothese verworfen wurde, einen weiteren Heteroskedastietest (vgl. Kapitel 6.2.4). Oder man kann direkt einen der üblichen asymptotischen Tests mit einem heteroskedastischen Modell als Alternative verwenden.

6.2.2.1.2 Unbekannter Schwellenwert

Muß der Schwellenwert geschätzt werden, so tritt das oben erwähnte Problem auf, daß dieser Parameter unter der Nullhypothese, daß die Regressionsparameter in den beiden Zuständen (und gegebenenfalls die Störvarianzen) gleich sind, nicht identifiziert ist und deswegen die klassischen Tests nicht die üblichen asymptotischen χ^2-Verteilungen besitzen.[147]

Für empirische Untersuchungen können in einigen Fällen die Simulationsstudien von CHAN [1991] verwendet werden. Dieser Autor verwendet nicht die gewöhnliche LQ-Statistik

$$\lambda = T\log\left(\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2_{H_1}}\right),$$

[146] Deswegen sind Tests auf Regimewechsel in diesem Fall eigentlich keine "Linearitätstests" im engeren Sinne (die Alternative ist formal ebenfalls ein lineares Modell).

[147] Einige theoretische Ergebnisse über LQ-Tests der Nullhypothese eines autoregressiven Modells gegen die Alternative eines Zeitreihen-Schwellenmodells, die aber für empirische Untersuchungen nicht besonders hilfreich sind, finden sich in CHAN [1990], TONG [1990], CHAN und TONG [1990], S. 233 ff., HANSEN [1996] sowie in den oben erwähnten Arbeiten von CARRASCO [1997] und GARCIA [1998].

wobei $\hat{\sigma}^2_{H_0}$ und $\hat{\sigma}^2_{H_1}$ die ML-Schätzer der Fehlervarianzen unter der Nullhypothese und unter der Alternative sind, sondern die Statistik

$$\lambda' = T\left(\frac{\hat{\sigma}^2_{H_0}}{\hat{\sigma}^2_{H_1}} - 1\right).$$

CHAN [1991] berechnet kritische Werte für diese Statistik für SETAR-Modelle aus Simulationen mit einigen wenigen Parameterwerten. Da λ' eine streng monotone Funktion von λ ist, lassen sich aus diesen simulierten Werten auch kritische Werte für die gewöhnliche LQ-Statistik λ berechnen. Vergleicht man diese mit denen von regulären LQ-Tests (also mit den kritischen Werten $\chi^2(n)_\alpha$, wobei n die Anzahl der zusätzlichen freien Parameter des Alternativmodells *inklusive* des Schwellenwertes und α das gewählte Niveau des Tests darstellt), so läßt sich etwas vereinfachend sagen, daß die simulierten Werte den kritischen Werten einer $\chi^2(n+n')$-Verteilung entsprechen, wobei die "Korrektur der Freiheitsgrade" n' etwa 1-2 für niedrige Werte von n beträgt und für größere Werte von n anwächst (auf etwa 3-4 bei $n = 20$). Exakte Angaben über das tatsächliche Signifikanzniveau lassen sich mit dieser groben Regel natürlich nicht machen. In vielen Anwendungen, in denen beobachtete Werte der LQ-Statistik weit ober- oder unterhalb dieser Grenze liegen, kann man damit aber fundierte Aussagen über die Bedeutung eines Schwellenmodells im Vergleich zu einem linearen Modell machen.[148]

6.2.2.2 Mischungs- und Markov-Modelle

Formuliert man die Nullhypothese eines linearen Modells dadurch, daß die Regressionsparameter in den beiden Zuständen (und gegebenenfalls die Störvarianzen) gleich sind, so sind für Mischungsmodelle die Aufenthaltswahrscheinlichkeit p und für Markov-Modelle die Startwahrscheinlichkeit ρ und die Übergangswahrscheinlichkeiten p^{11} und p^{22} unter der Nullhypothese nicht identifiziert.[149] Damit sind die gängigen asymptotischen Tests auch hier nicht ohne Modifikation anwendbar.[150]

---

[148] Eine weitere Möglichkeit in diesem Zusammenhang stellen Tests linearer gegen STAR-Modelle dar, die verallgemeinerte Schwellenmodelle darstellen (vgl. Kapitel 2.1, besonders Fußnote 9). Es ist zu vermuten, daß diese Tests im Vergleich zu Tests mit unspezifizierter nicht linearer Alternative vergleichsweise große Macht gegen Schwellenmodelle besitzen. Vgl. hierzu LUUKKONEN, SAIKKONEN und TERÄSVIRTA [1988].

[149] Man kann die Nullhypothese für Mischungsmodelle auch durch $p = 1$ (oder $p = 0$) formulieren, so daß nur ein Zustand eintreten kann. Dann sind die Regressionsparameter des anderen Zustandes unter der Nullhypothese nicht identifiziert, und überdies wird die Nullhypothese nicht mehr durch einen Wert im Inneren des zulässigen Parameterbereichs beschrieben, wie es die übliche asymptotische Testtheorie erfordert, sondern durch einen Wert am Rand des zulässigen Parameterbereichs. Eine analoge Bemerkung gilt für Markov-Modelle.

[150] Obwohl das Problem der nicht identifizierten Parameter schon lange bekannt ist und in der Literatur wenigstens seit der Arbeit von DAVIES [1977] diskutiert wird, wird es in Anwendungen immer wieder ignoriert, etwa von KON [1984] oder BOOTHE und GLASSMAN [1987].

Obwohl diesem Problem in letzter Zeit verstärkte Aufmerksamkeit gewidmet worden ist,[151] liegt bisher kein überzeugender Ansatz für empirische Untersuchungen vor. Hier bieten sich verschiedene praktikable Möglichkeiten an. In einigen Fällen kann man auf die simulierten Werte von GARCIA [1998] zurückzugreifen, der u.a. die empirische Verteilung der LQ-Statistik bestimmt hat, wenn die Nullhypothese ein Modell ohne Regressoren bzw. mit autoregressiven Größen und die Alternative das entsprechende homo- bzw. heteroskedastische Markov-Modell mit zustandsabhängiger Konstante ist. Seine Schlußfolgerung für das homoskedastische Modell ohne Regressoren lautet

> *"... that these critical values for the likelihood ratio are considerably higher than the values of $\chi^2(1)$, the distribution of the LR test in the classical theory."*[152]

Dabei muß man aber beachten, daß Garcia die unter der Nullhypothese nicht identifizierten Übergangswahrscheinlichkeiten bei diesem Vergleich nicht mitzählt. Evt. hat Garcia auch, so wie es in der neueren Literatur zu diesem Themenkomplex öfter getan wird, außer den beiden freien Parametern für die Übergangswahrscheinlichkeiten noch einen weiteren freien Parameter für die Startwahrscheinlichkeit verwendet. Die Freiheit in der Wahl dieser Parameter unter der Alternative schlägt sich natürlich in einem höheren maximalen Wert der Likelihood-Funktion nieder. Deswegen sollte man zum Vergleich seiner simulierten mit den erwarteten Werten auch diese zusätzlichen Freiheitsgrade berücksichtigen (wie es in der Literatur auch üblich ist) und die simulierten Werte im Beispiel des Modells ohne autoregressive Größen denen einer $\chi^2(3)$- bzw. einer $\chi^2(4)$-Verteilung gegenüberstellen. Dann kommt man zu einem ähnlichen Ergebnis wie für Schwellenmodelle mit unbekanntem Schwellenwert: Die kritischen Werte von Garcia liegen etwas über denen einer $\chi^2(4)$-Verteilung, wobei man die Abweichung durch zusätzliche 1-2 Freiheitsgrade berücksichtigen kann.[153] Zu einem solchen Schluß kamen schon GOLDFELD und QUANDT [1976] in einer Simulationsstudie, bei der die Nullhypothese ein lineares Modell und die Alternative ein heteroskedastisches Mischungsmodell mit zwei exogenen Regressoren ist:

> *"... the appropriate χ^2 seems to involve somewhat larger degrees of freedom than would be justified on a priori considerations (5 or 6 versus 4)."*[154]

---

[151] HANSEN [1992] schlägt einen Test von linearen Modellen gegen Markov-Modelle vor, der aber schwer zu implementieren ist, vgl. das *Erratum* von HANSEN [1995]. GARCIA [1998] wendet das Verfahren, das HANSEN [1996] mit Blick auf Schwellenmodelle beschreibt, auf Markov-Modelle an und gewinnt so einige theoretische Ergebnisse über die Verteilung der LQ-Statistik.

[152] Vgl. GARCIA [1998], S. 16. Die Ergebnisse für die andere untersuchten Modelle sind ähnlich.

[153] Für die anderen in GARCIA [1998] betrachteten Modelle ist das Bild ähnlich.

[154] GOLDFELD und QUANDT [1976], S. 21. Diese Autoren zählen die unter der Nullhypothese nicht identifizierten Parameter mit.

Eine Alternative zu simulierten Werten stellt der Ansatz von DAVIES [1977, 1987] dar, der eine Näherungsformel für die kritischen Werte von LQ-Tests angibt, die allgemein gültig ist, falls *ein* (eindimensionaler) Parameter unter der Alternative nicht identifiziert ist. Sie lautet

$$P(LQ > M) \approx P(\chi^2(s) > M) + VM^{\frac{1}{2}(s-1)} e^{-\frac{1}{2}M} 2^{-\frac{1}{2}s} \Big/ \Gamma\left(\tfrac{1}{2}s\right) ,$$

wobei LQ den Wert der LQ-Statistik bezeichnet, s die Anzahl der linear unabhängigen Restriktionen von *identifizierten* Parametern, und V die totale Variation von $LQ(\theta)^{1/2}$ als Funktion des nicht identifizierten Parameters θ über den zulässigen Bereich $[a,b]$, also

$$V = \int_a^b \left| \frac{\partial LQ(\theta)^{1/2}}{\partial \theta} \right| d\theta \ .$$

Diese Näherungsformel, die in der Literatur auch als *"Davies' quick rule"* bezeichnet wird, läßt sich z.B. für Mischungsmodelle mit $\theta = p$ verwenden. BOLDIN [1990] argumentiert, daß $LQ(p)^{1/2}$ für Mischungsmodelle i.d.R. nur *ein* Maximum hat, so daß (da an den Rändern des zulässigen Parameterbereichs, also bei $p = 0$ und $p = 1$, $LQ(p) = 0$ gilt) V dem doppelten Wert des Maximums der Funktion entspricht, d.h. $V = 2M^{1/2}$, und damit

$$P(LQ > M) \approx P(\chi^2(s) > M) + M^{\frac{1}{2}s} e^{-\frac{1}{2}M} 2^{1-\frac{1}{2}s} \Big/ \Gamma\left(\tfrac{1}{2}s\right) .$$

Daraus lassen sich approximative kritische Werte für verschiedene Werte von s berechnen, die in Tabelle 6 wiedergegeben sind.

Anzahl der Restriktionen	Niveau		
s	10 %	5 %	1 %
1	6,25	7,81	11,34
2	9,26	10,95	14,72
3	11,71	13,50	17,45
4	13,91	15,79	19,89
5	15,95	17,89	22,16
6	17,90	19,89	24,28
7	19,74	21,81	26,32
8	21,54	23,67	28,30
9	23,30	25,48	30,20
10	25,01	27,24	32,06

Tabelle 6: Kritische Werte des LQ-Tests für homogene Mischungsmodelle mit s unabhängigen Restriktionen von identifizierten Parametern mit Hilfe der Approximation von DAVIES [1987].

Will man diese kritischen Werte mit den oben erwähnten simulierten kritischen Werten vergleichen, so muß man zu dem Wert von s noch einen Freiheitsgrad addieren, der dem (unter der Nullhypothese nicht identifizierten) Parameter p entspricht. Damit erhält man wiederum

das Resultat, daß man für kleine Werte von s kritische Werte aus einer $\chi^2(n)$-Verteilung mit einer Korrektur der Freiheitsgrade um 1-2 nach oben verwenden kann, die für große Werte von s anwächst.[155]

Dies ist natürlich nur eine sehr grobe Faustregel. Man muß aber berücksichtigen, daß die exakten kritischen Werte (die man wegen der relativ aufwendigen ML-Schätzung von Regimewechselmodellen nicht für jedes in Frage stehende Modell mit Hilfe von Simulationen berechnen kann) nicht unerheblich von den wahren Parametern abhängen, wie die Untersuchungen in GARCIA [1998] zeigen. Die Variation der kritischen Werte für unterschiedliche Parameter liegt bereits in der Größenordnung des Fehlers, den man mit dieser Regel begehen wird.

In Anbetracht dieser (wenigstens für kleine bis mittlere Werte von s) übereinstimmenden Ergebnisse für Schwellenmodelle mit unbekanntem Schwellenwert und für homogene Mischungs- und Markov-Modelle ist davon auszugehen, daß die erwähnte Faustregel für die kritischen Werte auch für *inhomogene* Mischungs- und Markov-Modelle, die als Verallgemeinerungen der entsprechenden homogenen Modelle sowie von Schwellenmodellen aufgefaßt werden können, gelten wird.[156] Man sollte also ein lineares Modell zum Niveau α verwerfen, wenn die LQ-Statistik größer als $\chi^2(n+n')_\alpha$ ist, wobei n gleich der Anzahl aller (linear unabhängigen) Restriktionen von identifizierten und nicht identifizierten Parametern und n' der oben erwähnte Korrekturterm ist.

6.2.3 Tests zwischen Regimewechselmodellen mit verschiedener Dynamik des Zustandsprozesses

Um sich einen Überblick über die denkbaren Hypothesen, die sich auf die Zustandsprozesse von Regimewechselmodellen beziehen, zu verschaffen, ist ein Blick auf Abbildung 3 (S. 15) hilfreich, in der erläutert wird, welche Modelle sich durch Einschränkung aus anderen ergeben. Einige dieser Restriktionen, jedoch nicht alle, sind gewöhnliche explizite Restriktionen von Parametern, bei denen man (wie in Kapitel 6.1 dargelegt) davon ausgehen kann, daß die ML-Schätzer der betreffenden Parameter asymptotisch normalverteilt sind, so daß sich diese Restriktionen mit den klassischen Tests überprüfen lassen.[157] In anderen Fällen liegt diese asymptotische Normalverteilung nicht vor, und die klassischen Tests müssen modifiziert wer-

[155] Die Korrektur wächst mit dieser Approximationsformel allerdings auf 5-6 für $s = 10$ an. Die oben erwähnten Simulationsergebnisse zeigen aber, daß dies wohl auf mangelnde Präzision der Approximation zurückzuführen ist.

[156] Hier werden natürlich unter Umständen mehr (bei einem skalaren Regimeindikator bis zu 5) freie Parameter für die Modellierung des Zustandsprozesses verwendet, die unter der Nullhypothese nicht identifiziert sind.

[157] D.h. daß die üblichen Teststatistiken asymptotisch χ^2-verteilt sind.

den. In den folgenden Kapiteln werden diese Restriktionen und eventuelle Abweichungen von den entsprechenden χ^2-Verteilungen untersucht.

6.2.3.1 Tests zwischen verschiedenen Typen von Schwellenmodellen

Einen interessanten Fall stellen Tests von Hypothesen über den Schwellenwert eines Schwellenmodells dar, dessen ML-Schätzer ja *nicht* asymptotisch normalverteilt ist (vgl. Kapitel 6.1.1.2). Die Nullhypothese stellt hier also ein Schwellenmodell mit bekanntem, die Alternative eines mit unbekanntem Schwellenwert dar. Eine Vorstellung von der Verteilung der LQ-Teststatistik in dieser Situation vermittelt Tabelle 7, in der einige kritische Werte aus einer Simulation für diesen Fall wiedergegeben sind.

Niveau	10 %	5 %	1 %
kritischer Wert	3,55	5,14	9,33

Tabelle 7: Kritische Werte der LQ-Statistik aus 10.000 Realisationen einer Simulation unter der Nullhypothese eines Schwellenmodells mit bekanntem und der Alternative eines Schwellenmodells mit unbekanntem Schwellenwert
Zugrundeliegendes Modell: Keine Regressoren, homoskedastisch, Zeitreihenlänge 100 Beobachtungen, Mittelwerte = 1,0/-1,0, Störvarianz = 1,0, Schwellenvariable unabhängig identisch standardnormalverteilt, Schwellenwert = 0,0

Wenn man diese Werte mit denen einer $\chi^2(1)$-Verteilung vergleicht (da die Alternative ja einen zusätzlichen freien Parameter besitzt, den Schwellenwert), so erkennt man deutliche Abweichungen: Der kritische Wert zum 10 %-Niveau liegt etwa in der Mitte zwischen dem einer $\chi^2(1)$- und einer $\chi^2(2)$-Verteilung, der zum 1 %-Niveau ist etwas größer als der einer $\chi^2(2)$-Verteilung und der zum 5 %-Niveau etwas kleiner. Die Prozentpunkte der simulierten Verteilung sind also größer als die einer $\chi^2(1)$-Verteilung, und ihre Form ist anders.[158]

6.2.3.2 Tests zwischen verschiedenen Typen von Mischungs- und Markov-Modellen

Üblicherweise werden die Aufenthalts- bzw. Übergangswahrscheinlichkeiten inhomogener Mischungs- und Markov-Modelle durch logistische Funktionen dargestellt, also etwa für Mischungsmodelle

$$p(z_t) = \frac{e^{\delta+\gamma z_t}}{1+e^{\delta+\gamma z_t}} .$$

Für inhomogene Markov-Modelle gibt es entsprechend 4 freie Parameter δ_1 , δ_2 , γ_1 und γ_2. Die Nullhypothese eines *homogenen* gegen die Alternative eines *inhomogenen* Mischungs-

[158] Da *verallgemeinerte* Schwellenmodelle in der empirischen Literatur keine nennenswerte Rolle spielen, wird auf sie an dieser Stelle nicht näher eingegangen.

(bzw. Markov-)Modells läßt sich durch die reguläre Restriktion $\gamma = 0$ (bzw. $\gamma_1 = \gamma_2 = 0$) ausdrücken, so daß $p = e^{\delta}/1 + e^{\delta}$ die konstante Aufenthaltswahrscheinlichkeit in Zustand 1 darstellt (analog erhält man für Markov-Modelle die konstanten Übergangswahrscheinlichkeiten). Die Teststatistiken sollten $\chi^2(1)$- bzw. $\chi^2(2)$-verteilt sein.[159] Bei diesem Test wird untersucht, ob der Einfluß des Regimeindikators auf die Aufenthalts- bzw. die Übergangswahrscheinlichkeiten signifikant ist.

Die klassischen Tests sind vermutlich ebenso gültig für den Test von homogenen oder inhomogenen Mischungsmodellen gegen die entsprechenden Markov-Modelle. Die Teststatistiken sollten, wenn man die Startwahrscheinlichkeit im Zustand 1 als freien Parameter verwendet, im ersten Fall $\chi^2(2)$-verteilt und im zweiten Fall $\chi^2(3)$-verteilt sein.[160] Falls man annimmt, daß die Markov-Kette stationär ist, reduziert sich die Anzahl der Freiheitsgrade in beiden Fällen um 1. Hier steht die Frage im Vordergrund, ob Autokorrelation der Zustände, die sich in längeren Folgen ein und desselben Zustandes ausdrückt, signifikant ist, oder ob die Unabhängigkeit der Zustände mit den Daten vereinbar ist.

6.2.3.3 Tests von Schwellen- gegen inhomogene Mischungs- und Markov-Modelle

Wieder anders ist die Situation beim Test von Schwellenmodellen gegen inhomogene Mischungs- und Markov-Modelle, bei dem untersucht wird, ob das Über- und Unterschreiten des Schwellenwertes durch den Regimeindikator den Wechsel zwischen den Zuständen erklären kann, oder ob eine zusätzliche stochastische Komponente die Anpassung des Modells deutlich erhöht. Für diese Situationen sind andere Parametrisierungen als die oben erwähnte geeigneter, etwa (für inhomogene Mischungsmodelle)

$$p(z_t) = \frac{e^{\gamma(z_t - c)}}{1 + e^{\gamma(z_t - c)}} \; .$$

Im Grenzfall $\gamma \to \infty$ konvergiert diese Funktion gegen eine Stufenfunktion mit einem Sprung bei c, und damit geht das inhomogene Mischungsmodell in ein Schwellenmodell mit Schwellenwert c über. Ähnliches gilt für inhomogene Markov-Modelle, die (bei entsprechender Parametrisierung) im Grenzfall $\gamma_1 \to \infty$, $\gamma_2 \to \infty$ in verallgemeinerte Schwellenmodelle übergehen.[161] Diese Testsituationen sind bisher in der Literatur noch nicht untersucht worden.

---

[159] Dies wird z.B. bei FILARDO [1994], S. 302 f., verwendet.

[160] Dies nutzen z.B. COSSLETT und LEE [1985], S. 83 ff. – Hier ließe sich auch ein White-Test verwenden, vgl. HAMILTON [1996], S. 139 f.

[161] Vgl. hierzu auch Kapitel 2.2.2.

Auf solche Restriktionen läßt sich die asymptotische Testtheorie nicht ohne weiteres anwenden. Dennoch könnte man vermuten, daß der Satz über die asymptotische χ^2-Verteilung der üblichen Teststatistiken in diesem Grenzfall weiterhin gültig ist, gilt er doch für jedes beliebig große *endliche* γ. Dem ist aber entgegenzuhalten, daß andere vergleichbare Situationen wohlbekannt sind, in denen die Nullhypothese einem „Randfall" eines parametrischen Modells entspricht, das – wiewohl „beliebig nahe" an der Alternative – eine gänzlich andere stochastische Struktur als die Alternativmodelle hat, so daß als Konsequenz dieser Strukturveränderung die üblichen Teststatistiken eine andere Verteilung als unter jeder anderen in Frage kommenden Nullhypothese haben. Das bekannteste Beispiel hierfür sind wohl Einwurzeltests: Unter der Nullhypothese einer Einheitswurzel ist ein Zeitreihenmodell nicht stationär, und die t-Statistik aus der Hilfsregression eines Dickey-Fuller-Tests ist nicht mehr t-verteilt (wie es unter der Alternative eines stationären Modells ohne Einheitswurzel der Fall wäre), sondern Dickey-Fuller-verteilt.[162] Weitere Beispiele für solche Situationen (in denen die Verteilung der üblichen Teststatistiken unbekannt ist) sind der Test der Nullhypothese einer Normalverteilung gegen die Alternative einer t-Verteilung,[163] oder der eines Schwellenmodells gegen die Alternative eines STAR-Modells.[164]

Es ist daher nicht unwahrscheinlich, daß die Verteilung der üblichen Likelihood-orientierten Teststatistiken in dem betrachteten Fall auch asymptotisch von der erwarteten χ^2-Verteilung abweicht. Aus diesem Grund wurden wieder einige Simulationsstudien durchgeführt, in denen die Verteilung der LQ-Statistik unter der Nullhypothese eines Schwellenmodells und der Alternative eines inhomogenen Mischungs- bzw. Markov-Modells untersucht wurde. Die Ergebnisse sind in Tabelle 8 dargestellt.

Diese kritischen Werte wären zu vergleichen mit denen einer $\chi^2(2)$- bzw. einer $\chi^2(4)$- Verteilung. Die Werte aus Simulation 1 liegen für das 5 %- und das 10 %-Niveau leicht unter denen der erwarteten $\chi^2(2)$-Verteilung, und für das 1 %-Niveau darüber. Dies kann auf eine prinzipielle Abweichung der hier untersuchten von der in Frage kommenden theoretischen Verteilung zurückzuführen sein, auf mangelnde Asymptotik aufgrund der Zeitreihenlänge von 100 Beobachtungen, oder auf die Zahl von Simulationen, die aufgrund der aufwendigen

[162] Vgl. hierzu DICKEY und FULLER [1981], RÜDEL [1989] oder BANERJEE u.a. [1993].

[163] Hierfür wird in BOOTHE und GLASSMAN ohne weitere Rechtfertigung der gewöhnliche LQ-Test verwendet.

[164] Zu STAR-Modellen vgl. Fußnote 9. – Teräsvirta vermutet, daß die asymptotische χ^2-Verteilung der LQ-Statistik auch hier gilt, kann dies aber nicht beweisen (private Mitteilung). Eine Lösung des oben beschriebenen Testproblems für diese Situation würde es also auch gestatten, das Schwellenmodell des US-BSP von POTTER [1995] gegen das STAR-Modell von GRANGER und TERÄSVIRTA [1993] zu testen, also eine Antwort auf die Frage zu geben, ob ein Modell mit abrupten Sprüngen zwischen den Phasen des Konjunkturzyklusses mit den Daten verträglich ist, oder ob der "glatte" Übergang der STAR-Modelle signifikant ist.

Schätzverfahren unter der Alternative in diesem Fall relativ gering ist. Einheitlicher ist das Bild für Simulation 2: Die simulierten kritischen Werte liegen hier durchgängig zwischen denen einer $\chi^2(3)$- und der erwarteten $\chi^2(4)$-Verteilung.

Niveau		10 %	5 %	1 %
kritischer Wert	Simulation 1	3,89	5,41	9,76
	Simulation 2	6,52	8,83	11,54

Tabelle 8: Kritische Werte des LQ-Tests aus 1.000 Realisationen einer Simulation unter der Nullhypothese eines Schwellenmodells und der Alternative eines Mischungs- und eines Markov-Modells Zugrundeliegende Modelle: Nullhypothese wie in Tabelle 7 (100 Beobachtungen,), Alternative: Simulation 1: Inhomogenes Mischungsmodell, Simulation 2: Inhomogenes, stationäres Markov-Modell

Zusammenfassend läßt sich sagen, daß die simulierten Werte durchaus in der Größenordnung derjenigen der erwarteten χ^2-Verteilungen liegen. Um kritische Werte zu erhalten, die genaue Aussagen über das Niveau von LQ-Tests für diese Situation gestatten, wären selbstverständlich wesentlich umfangreichere Simulationen als die hier präsentierten erforderlich. Dennoch können diese ersten Ergebnisse als Richtschnur in empirischen Untersuchungen verwendet werden, insbesondere, wenn die beobachteten Werte der LQ-Statistik deutlich größer oder kleiner als die hier angegebenen kritischen Werte sind.

6.2.4 Spezifikationstests

Ist man in einer empirischen Anwendung zu dem Schluß gekommen, daß ein bestimmter Typ von Regimewechselmodellen angebracht ist, so sollte man noch verschiedene Restriktionen für die Regressionsparameter testen. Beispiele für solche Hypothesen sind:

- Ein (oder mehrere) Regressionsparameter ist (sind) *nicht* zustandsabhängig.[165]
- Die Störvarianz ist nicht zustandsabhängig, d.h. das Modell ist homoskedastisch.
- Ein Regressor hat nur in einem der beiden Zustände Einfluß auf die abhängige Variable, d.h. der Wert des zugehörigen Parameters in dem entsprechenden Zustand ist Null.

Ist die Aufteilung der Daten auf die Zustände bekannt, so kann man exakte Tests für diese Probleme angeben. Auf diese Situation wird in Kapitel 6.2.4.1.1 eingegangen. In den übrigen Fällen, die in den darauf folgenden Kapiteln 6.2.4.1.2 und 6.2.4.2 behandelt werden, ist man auf asymptotische Tests angewiesen.

[165] Diese wichtige Hypothese wird in der Literatur fast völlig vernachlässigt. Eine Ausnahme bilden die Untersuchungen in LÜDEKE [1973].

6.2.4.1 Schwellenmodelle

6.2.4.1.1 Bekannter Schwellenwert

Da diese Modelle linear sind (vgl. Kapitel 4.2.1), können *lineare Hypothesen über die Regressionsparameter* im homoskedastischen Fall mit exogenen Regressoren mit dem üblichen F-Test getestet werden (und für Hypothesen, die einzelne Parameter betreffen, ist der t-Test anwendbar). Für heteroskedastische Modelle muß man auf die üblichen asymptotischen Tests zurückgreifen.[166]

Diese kann man auch für einen *Test auf Heteroskedastie* anwenden. Eine besondere Situation tritt hierbei in dem häufig betrachtet Fall auf, daß *alle* Regressionsparameter regimeabhängig sind. Dann gilt (mit den Bezeichnungen aus Kapitel 4.2.1.2, Fall **(2)** und **(3)**) für die ML-Schätzer der Varianzen $\hat{\sigma}^2$ bzw. $\hat{\sigma}_1^2$ und $\hat{\sigma}_2^2$

$$\hat{\sigma}^2 = \frac{T_1\hat{\sigma}_1^2 + T_2\hat{\sigma}_2^2}{T} .$$

Deswegen berechnet sich die LQ-Statistik aus den auf S. 43 angegebenen Werten für das Maximum der Log-Likelihood zu

$$\begin{aligned}\lambda &= 2\left(-\frac{T_1}{2}\log\left(\hat{\sigma}_1^2\right)-\frac{T_2}{2}\log\left(\hat{\sigma}_2^2\right)+\frac{T}{2}\log\left(\hat{\sigma}^2\right)\right)\\ &= 2\log\left(\frac{\left(T_1\hat{\sigma}_1 + T_2\hat{\sigma}_2\right)^{T_1+T_2}}{\hat{\sigma}_1^{T_1}\hat{\sigma}_2^{T_2}}\right)-2T\log(T)\\ &= 2\log\left((T_1+T_2/v)^{T_1}(T_1v+T_2)^{T_2}\right)-2T\log(T)\end{aligned}$$

mit $v = \hat{\sigma}_1^2/\hat{\sigma}_2^2$. Diese LQ-Statistik ist, wie man an ihren ersten beiden Ableitungen nach v sieht, minimal für $v = 1$ (d.h. $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$), streng monoton fallend für $v < 1$ ($\hat{\sigma}_1^2 < \hat{\sigma}_2^2$) und streng monoton wachsend für $v > 1$ ($\hat{\sigma}_1^2 > \hat{\sigma}_2^2$). Der *exakte* LQ-Test ist hier also äquivalent zu einem F-Test. Die kritischen Bereiche zum Niveau α für die *einseitigen* Alternativen $\sigma_1^2 < \sigma_2^2$ bzw. $\sigma_1^2 > \sigma_2^2$ sind

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \geq F_{\alpha,T_1-1,T_2-1} \text{ bzw. } \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \geq F_{\alpha,T_2-1,T_1-1}$$

und für die *zweiseitige* Alternative $\sigma_1^2 \neq \sigma_2^2$

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \geq F_{\alpha/2,T_1-1,T_2-1} \text{ , falls } \hat{\sigma}_1^2 \geq \hat{\sigma}_2^2 ,$$

[166] Vgl. hierzu auch JUDGE u.a [1985], Kapitel 5.7.1.

und

$$\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \geq F_{\alpha/2,T_2-1,T_1-1} \text{ , falls } \hat{\sigma}_1^2 < \hat{\sigma}_2^2 \text{ .}$$

6.2.4.1.2 Unbekannter Schwellenwert

In Kapitel 6.1 wurde erläutert, daß bei unbekanntem Schwellenwert unter bestimmten Voraussetzungen die ML-Schätzer für die Regressionsparameter und die Varianzen asymptotisch normalverteilt sind, mit derselben Varianz-Kovarianz-Matrix wie im Falle eines bekannten Schwellenwertes. Deshalb ist zu vermuten, daß auch die asymptotischen Tests von Hypothesen, die sich auf diese Parameter beziehen, weiter Gültigkeit haben (da für ihre Herleitung ja eben diese asymptotische Normalverteilung verwendet wird).
Test von Hypothesen über den Schwellenwert wurden bereits in Kapitel 6.2.3.1 behandelt.

6.2.4.2 Mischungs- und Markov-Modelle

Wie in Kapitel 6.1.2 dargelegt, ist die asymptotische Normalverteilung der ML-Schätzer nur für einige wenige Fälle wirklich bewiesen. In den anderen Fällen wird sie allgemein als gegeben angenommen. Unter der Annahme der Gültigkeit dieser asymptotischen Normalverteilung kann man wiederum die asymptotischen Tests von Hypothesen, die sich auf diese Parameter beziehen, anwenden.

6.2.5 Tests zwischen getrennten Hypothesen

In empirischen Anwendungen kann es von Interesse sein, eine Auswahl zwischen zwei Modellen zu treffen, die *getrennt* (*separate* oder *nonnested*) sind, von denen sich also keines durch Parameterrestriktionen aus dem jeweils anderen ergibt. So kann man etwa vor der Frage stehen, ob ein homogenes Markov-Modell oder ein Schwellenmodell angebracht ist;[167] oder, welche von zwei konkurrierenden Einflußgrößen eine überlegene Erklärungsgüte für die abhängige Variable besitzt.

Die ersten systematischen Untersuchungen zu dem Problem von Tests zwischen getrennten Hypothesen stammen von COX [1961, 1962], der Verallgemeinerungen des klassischen LQ-Tests auf diese Situation vorschlug. Seitdem sind in der ökonometrischen Literatur zahlreiche weitere Testverfahren für solche Situationen vorgestellt worden,[168] von denen die Tests zwi-

[167] Ein Beispiel für diese Situation wäre ein Test zwischen dem Schwellenmodell des US-BSP von POTTER [1995] und dem Markov-Modell von HAMILTON [1989] für dieselben Daten.

[168] Überblicksdarstellungen zu diesem Problemkreis finden sich z.B. in DAVIDSON und MCKINNON [1982], GOURIEROUX, MONFORT und TROGNON [1983], MCALEER [1995] und GOURIEROUX und MONFORT [1995b], Kapitel 22.2.

schen getrennten klassischen Regressionsmodellen (d.h. solchen, von denen jedes mindestens eine Einflußgröße enthält, die in dem jeweils anderen fehlt) mittlerweile in vielen Ökonometrie-Lehrbüchern dargestellt werden.[169]

Naheliegend bei dem Problem eines solchen Tests zwischen zwei getrennten Modellen M_0 und M_1 ist der folgende, in vielen Lehrbüchern empfohlene Ansatz:[170] Es ist nach einem *umfassenden* Modell M_2 zu suchen, das die beiden in Frage stehenden Modelle M_0 und M_1 als Spezialfälle enthält. Es werden zwei konventionelle Tests der Nullhypothesen M_0 und M_1 gegen die Alternative dieses umfassenden Modells M_2 durchgeführt. Man testet also nicht direkt M_0 gegen M_1 und umgekehrt, sondern jedes der beiden Modelle gegen ein weiteres, das zusätzlich zu den Charakteristika des als Nullhypothese verwendeten Modells noch die des jeweils anderen enthält. Wird die Nullhypothese in einem dieser beiden Tests verworfen, so interpretiert man dies dahingehend, daß diese zusätzlichen Modellstrukturen wesentliche Eigenschaften der Daten beschreiben und deswegen in der Modellierung berücksichtigt werden sollten.

Die vier möglichen Ergebniskombinationen der beiden konventionellen Tests führen demnach zu folgender Entscheidungsregel:

	M_0 verworfen	M_0 nicht verworfen
M_1 verworfen	M_2 ist M_0 und M_1 vorzuziehen	M_0 ist M_1 und M_2 vorzuziehen
M_1 nicht verworfen	M_1 ist M_0 und M_2 vorzuziehen	keine Entscheidung möglich

Tabelle 9: Entscheidungsregel des Tests zwischen getrennten Hypothesen je nach Ausgang der Tests von M_0 (Vorzeile) und M_1 (Vorspalte), jeweils gegen M_2

Das Verfahren liefert keine Entscheidung, falls keine der beiden Restriktionen verworfen werden kann, und stellt deswegen auch keinen Test im herkömmlichen Sinne dar. Diese Schwierigkeit teilt das Verfahren jedoch mit allen Tests getrennter Hypothesen, im Unterschied zu reinen Diskriminationsverfahren, die zum Ziel haben, das in gewisser Hinsicht beste Modell aus den zur Diskussion stehenden auszuwählen (wie z.B. mit Hilfe der Informationskriterien). Auch ist nicht klar, wie sich die α- und β-Fehler der Einzeltests zur gesamten Irrtumswahrscheinlichkeit verhalten.

Ein entscheidender Vorteil dieses Verfahrens gegenüber den Alternativen ist, daß es einen interpretierbaren Schluß zuläßt, falls die Nullhypothese in beiden Fällen verworfen wird: M_0

[169] Vgl. z.B. GREENE [1997], Kapitel 7.10, oder JUDGE u.a. [1985], Kapitel 21.3.

[170] Vgl. z.B. HARVEY [1993], S. 177 ff., oder GOURIEROUX und MONFORT [1995b], Kapitel 22.2.7.

und M_1 beschreiben die Situation nur unvollständig, das umfassende Modell ist beiden vorzuziehen.

Die Auswahl eines solchen umfassenden Modells ist mit einer gewissen Willkür behaftet. Sehr allgemeine Ansätze sind die von Quandt bzw. Atkinson, bei denen die Likelihood-Funktion des umfassenden Modells als gewichtetes arithmetisches bzw. geometrisches Mittel der Likelihood-Funktionen der Einzelmodelle konstruiert wird.[171] Diese "Hybrid-Modelle" sind aber in den meisten Fällen reine stochastische Konstrukte und lassen sich dann i.d.R. kaum noch substanzwissenschaftlich deuten.[172]

Sind die Modelle, zwischen denen getestet werden soll, verwandt, so bietet sich dagegen in manchen Situationen ein umfassendes Modell an, dessen Struktur und Parameter inhaltliche Bedeutung besitzen, und das gegebenenfalls auch – falls die Testergebnisse nahelegen, daß es M_0 und M_1 vorzuziehen ist – zu weiteren Test- und Prognosezwecken herangezogen werden kann.

Aus Abbildung 3 (S. 15) ist ersichtlich, wie nach diesem Verfahren verschiedene Regimewechselmodelle gegeneinander getestet werden können: Da inhomogene Markov-Modelle alle anderen hier betrachteten Modelle umfassen, gibt es immer wenigstens dieses eine interpretierbare umfassende Modell, das für Tests zwischen getrennten Hypothesen verwendet werden kann.

Einen besonders interessanten Fall stellt in diesem Zusammenhang das Problem eines Tests zwischen Markov- und Schwellenmodellen dar. Insbesondere die Zeitreihenversionen dieser Modelle stehen in der ökonometrischen Literatur relativ gleichberechtigt nebeneinander, und ein Test zwischen ihnen wäre zur Beurteilung der konkurrierenden Modelle hilfreich. Dieses Problem wurde bisher in der Literatur weitestgehend vernachlässigt. Die einzige Ausnahme hierzu bildet die Arbeit von CLEMENTS und KROLZIG [1998], die einen Vergleich zwischen Markov- und Schwellenmodellen für das US-BSP anhand ihrer Prognosegüte anstellen. Das Ergebnis solch eines Vergleichs hängt natürlich entscheidend von dem speziellen Verlauf der untersuchten Zeitreihe außerhalb des für die Parameterschätzung verwendeten Zeitraums ab, weswegen dieses Verfahren keinen Ersatz für einen statistischen Test darstellt.[173]

---

[171] Vgl. z.B. GOURIEROUX und MONFORT [1995b], S. 298.

[172] Ähnliches gilt für das Davidson-McKinnon-Modell, das getrennte Regressionsmodelle "umfaßt" und die Grundlage für den *J*- und den *P*-Test zwischen solchen Modellen bildet, vgl. GOURIEROUX und MONFORT [1995b], S. 301 ff.

[173] In CARRASCO [1997] werden außerdem *Wald-Encompassing-Tests* (vgl. MIZON und RICHARD [1986]) zwischen Schwellen- und Strukturbruch- sowie zwischen Markov- und Strukturbruchmodellen vorgeschlagen. Dieser Ansatz läßt sich aber nicht ohne weiteres auf den Fall eines Tests zwischen Schwellen- und Markov-Modellen ausdehnen, vgl. CARRASCO [1997], S. 3.

Es dürfte darum von Interesse sein, die oben erläuterte Methodik auf diesen Fall anzuwenden. Dafür benötigt man einen (herkömmlichen) Test zwischen homogenen und inhomogenen Markov-Modellen, der aber – wie in Kapitel 6.2.3 ausgeführt – unproblematisch ist, sowie einen zwischen Schwellen- und inhomogenen Markov-Modellen. In dieser Situation kann der verallgemeinerte LQ-Test aus Kapitel 6.2.3 verwendet werden, auf den an dieser Stelle deswegen genauer eingegangen wurde. Stehen ein Schwellen- und ein Mischungsmodell zur Diskussion, so kann als umfassendes Modell ein inhomogenes Mischungsmodell zusammen mit dem in Kapitel 6.2.3 untersuchten verallgemeinerten LQ-Test zwischen Schwellen- und inhomogenen Mischungsmodellen (sowie einem zwischen homogenen und inhomogenen Mischungsmodellen) herangezogen werden.

7 Tests von Regimewechselmodellen für Wochendaten des FF/DM-Wechselkurses

Die im letzten Kapitel erläuterten Tests im Zusammenhang mit Regimewechselmodellen sollen im folgenden auf die verschiedenen in Kapitel 5 geschätzten Modelle für den FF/DM-Wechselkurs angewendet werden.

7.1 Tests linearer Modelle gegen Regimewechselmodelle

Als erstes stellt sich die Frage nach der Signifikanz der geschätzten Modelle. Zu testen ist die Nullhypothese eines Modells ohne Regimewechsel gegen die Alternativen der verschiedenen Typen von Regimewechselmodellen, wofür hier die in 6.2.2 erläuterten LQ-Tests verwendet werden. Die Werte der LQ-Teststatistik sind in Tabelle 10 aufgeführt.

	Schwellenmodell	Mischungsmodell	Markov-Modell
Wert der LQ-Teststatistik	157,50	331,98	426,96

Tabelle 10: LQ-Teststatistiken für die Tests des linearen Modells gegen Regimewechselmodelle

Da hier der (unter der Nullhypothese nicht identifizierte) Schwellenwert des Schwellenmodells geschätzt wird, folgt die LQ-Teststatistik in diesem Fall nicht der üblichen $\chi^2(n)$-Verteilung, wobei n die Anzahl der zusätzlichen freien Parameter des Alternativmodells inklusive des Schwellenwertes darstellt, d.h. hier $n = 3$. Folgt man der in Kapitel 6.2.2.1.2 angegebenen Faustregel, daß man statt dessen die kritischen Werte einer $\chi^2(n+n')$-Verteilung verwenden muß, wobei die "Korrektur der Freiheitsgrade" n' etwa 1-2 für den hier verwendeten niedrigen Wert von n beträgt, so erhält man für $n' = 2$ zum 1 %-Niveau einen kritischen Wert von 15,09 und damit das Ergebnis, daß die Nullhypothese verworfen wird.

Das Problem von unter der Nullhypothese nicht identifizierten Parametern und daraus resultierenden exotischen Verteilungen der LQ-Teststatistiken besteht auch für Mischungs- und Markov-Modelle. Für das Mischungsmodell kann man die kritischen Werte aus Tabelle 6 (S. 100) verwenden, die aus der Davies-Approximation berechnet wurden. Der entsprechende kritische Wert zum 1 %-Niveau für den hier untersuchten Fall mit $s = 2$ beträgt 14,72. Die Nullhypothese wird also auch hier verworfen.

Für das Markov-Modell schließlich sind die simulierten Werte von GARCIA [1998] anwendbar. Der kritische Wert zum 1 %-Niveau für dieses Modell lautet 17,52. Wieder wird die Nullhypothese verworfen.

Zusammenfassend erhält man das Ergebnis, daß die Tests keinen Zweifel daran lassen, daß *alle* drei untersuchten Regimewechselmodelle die Daten signifikant besser beschreiben als das lineare Modell. Damit stellt sich die Frage nach dem „besten" unter den betrachteten Regimewechselmodellen.

7.2 Tests zwischen verschiedenen Typen von Regimewechselmodellen

Relativ einfach ist die Situation für den Vergleich zwischen dem Mischungs- und dem Markov-Modell. Das Mischungsmodell ist eine Restriktion des Markov-Modells, und man kann davon ausgehen, daß die LQ-Teststatistik für den Test der Nullhypothese des Mischungsmodells gegen die Alternative des Markov-Modells $\chi^2(1)$-verteilt ist (da hier kein separater freier Parameter für die Startwahrscheinlichkeit der Markov-Kette geschätzt wird). Der Wert der Teststatistik beträgt 94,98 bei einem kritischen Wert von 6,64 zum 1 %-Niveau Damit wird die Nullhypothese verworfen: Die Häufung von Beobachtungen aus den beiden Zuständen, die nicht durch das Mischungsmodell, wohl aber durch das Markov-Modell erfaßt wird, stellt ein wesentliches Charakteristikum der Daten dar.

Dieses *stylised fact* wird aber auch von dem Schwellenmodell beschrieben. Welches von beiden ist dem anderen vorzuziehen? Da die beiden Modelle getrennt sind, läßt sich kein herkömmlicher Test zwischen ihnen durchführen. Eine Möglichkeit ist das in Kapitel 6.2.5 erläuterte Vorgehen, das darin besteht, beide in Frage stehenden Modelle gegen ein umfassendes Modell zu testen. Dafür bietet sich in dieser Situation, wie schon in Kapitel 6.2.5 erläutert, ein inhomogenes Markov-Modell an.

Die ML-Schätzwerte eines solchen Modells, dessen Übergangswahrscheinlichkeiten in der Form

$$p^{ii}(z_t) = \frac{e^{\delta_i + \gamma_i z_t}}{1 + e^{\delta_i + \gamma_i z_t}}, \; i = 1, 2$$

von der Zinsdifferenz z_t abhängen, sind in Tabelle 11 wiedergegeben.[174] Ein Vergleich mit Tabelle 5 (S. 79) zeigt, daß die Schätzergebnisse für die Mittelwerte und die Standardabweichungen nur unwesentlich von denen des homogenen Modells abweichen. Der geschätzte Koeffizient $\hat{\gamma}_2$ hat das erwartete Vorzeichen: Bei wachsenden Zinsdifferenzen sinkt die Wahrscheinlichkeit für das Verbleiben in dem Zustand mit der niedrigen Volatilität. Das ne-

[174] Zu dem verwendeten Schätzverfahren vgl. den Anhang von Kapitel 5. Da die Normalgleichungen für die Parameter des Zustandsprozesses nicht linear sind (vgl. Kapitel 4.3.2.2.4), wurde für den M-Schritt des EM-Algorithmus die im Anhang B von Kapitel 4 erläuterte Näherungsformel verwendet.

gative Vorzeichen von $\hat{\gamma}_1$ ist jedoch nicht das erwartete. Dieser Koeffizient hat außerdem einen sehr hohen geschätzten Standardfehler.

$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
1,04 (0,29)	1,80 (0,18)	-0,04 (0,10)	-0,10 (0,09)	0,028 (0,052)	-0,0077 (0,0072)	0,608 (0,093)	0,141 (0,016)
Maximaler Wert der Log-Likelihood: 44,80							

Tabelle 11: Schätzwerte eines inhomogenen Markov-Modells

Zur Illustration ist der Verlauf der Übergangswahrscheinlichkeiten für dieses Modell in Abbildung 11 dargestellt. Die Übergangswahrscheinlichkeiten des inhomogenen Modells oszillieren (je nach Zinsentwicklung) in mäßigem Umfang um die (konstanten) des homogenen.

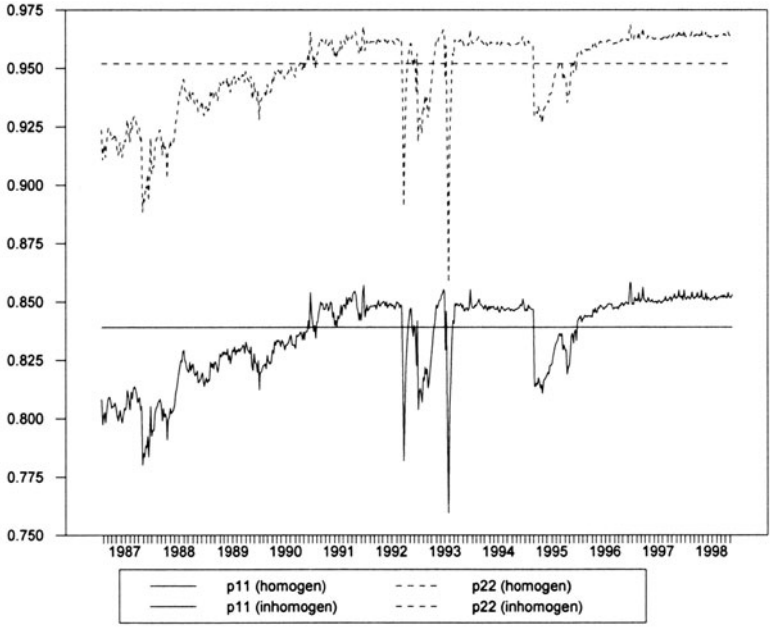

Abbildung 11: Vergleich der Übergangswahrscheinlichkeiten des homogenen und des inhomogenen Markov-Modells

Insgesamt sind die beiden Modelle sehr ähnlich. Dies zeigt sich insbesondere an einem LQ-Test der Nullhypothese des homogenen gegen die Alternative des inhomogenen Modells. Die Teststatistik, die unter der Nullhypothese $\chi^2(2)$-verteilt ist, hat den Wert 1,36 mit einem

p-Wert von 0,51. Die Nullhypothese wird also nicht verworfen: Der zusätzliche Erklärungsgehalt des inhomogenen Modells ist zu gering.

Will man die Nullhypothese des Schwellenmodells gegen die Alternative des inhomogenen Markov-Modells testen, so kann man auf die Ergebnisse aus Kapitel 6.2.3.3 zurückgreifen. Da davon auszugehen ist, daß die kritischen Werte in der Größenordnung einer $\chi^2(3)$ – Verteilung liegen,[175] ist das Testergebnis hier eindeutig: Bei einem Wert der Teststatistik von 270,82 wird die Nullhypothese deutlich verworfen.

Das Ergebnis des in Kapitel 6.2.5 beschriebenen Tests zwischen den getrennten Hypothesen eines homogenen Markov- und eines Schwellenmodells lautet also: Das Markov-Modell ist dem Schwellenmodell vorzuziehen. Damit ist aus den insgesamt 4 untersuchten Regimewechselmodellen das mit dem höchsten Erklärungsgehalt bei sparsamer Verwendung von Parametern ausgewählt worden.

7.3 Spezifikationstests

Zuletzt soll noch das Markov-Modell, welches im letzten Kapitel aus den geschätzten Modellen ausgewählt wurde, weiter untersucht werden. Dabei soll zum einen geprüft werden, ob der auffällige Unterschied zwischen den Volatilitäten in den beiden Zuständen signifikant ist. Dies könnte mit Hilfe eines LQ-Tests untersucht werden. Hier bietet sich aber ein Wald-Test an, der den Vorteil hat, daß das Modell nicht unter der Nullhypothese der Homoskedastie geschätzt werden muß. Der Wert der Teststatistik ist 37,94. Der Vergleich mit den kritischen Werten der $\chi^2(1)$-Verteilung zeigt, daß die Nullhypothese zu allen üblichen Niveaus verworfen wird. Die beobachtete Heteroskedastie ist also signifikant.

Zum anderen soll getestet werden, ob die Mittelwerte tatsächlich zustandsabhängig sind. Der Wert der Wald-Teststatistik der Nullhypothese eines Modells mit zustands*un*abhängigem Mittelwert ist 0,46 mit einem p-Wert von 0,50. Folglich wird die Nullhypothese nicht verworfen, die Mittelwerte sind nicht signifikant zustandsabhängig. Die Schätzwerte eines Modell mit der Restriktion gleicher Mittelwerte sind in Tabelle 12 dargestellt. Daraus läßt sich nun auch der Wert der LQ-Statistik für den Test der Nullhypothese eines zustandsunabhängigen Mittelwertes berechnen. Ihr Wert (0,52) liegt nahe bei dem Wert der zuletzt angegebenen Wald-Teststatistik und führt zu demselben Testergebnis.

Die Übergangswahrscheinlichkeiten und die Standardabweichungen der geschätzten Parameter sind fast identisch mit denen des Modells mit verschiedenen Mittelwerten in den Zu-

ständen. Auffällig sind die niedrigen Werte der geschätzten Standardfehler dieser Parameter, die im Falle der Übergangswahrscheinlichkeiten etwas kleiner als die des Modells mit verschiedenen Mittelwerten und im Falle der Standardabweichungen der Störterme deutlich niedriger sind.

$\hat{p}^{11}$	$\hat{p}^{22}$	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,841 (0,041)	0,953 (0,012)	-0,0070 (0,0068)	0,612 (0,028)	0,142 (0,005)	43,86

Tabelle 12: Geschätzte Parameter eines Markov-Modells mit zustandsunabhängigem Mittelwert

Der Erwartungswert der Verweildauer beträgt 6,3 Wochen für den Zustand mit der hohen und 21,3 Wochen für den Zustand mit der niedrigen Volatilität. Die unbedingte Wahrscheinlichkeit für den volatilen Zustand ist wieder 0,228. Die geglätteten Wahrscheinlichkeiten dieses Modells sind denen des Modells mit verschiedenen Mittelwerten (vgl. Abbildung 10, S. 80) so ähnlich, daß auf ihre Darstellung hier verzichtet wird.

Der Mittelwert, d.h. der gemeinsame Drift in den beiden Zuständen, ist, wie man an dem geschätzten Standardfehler in Tabelle 12 sieht, nicht signifikant von Null verschieden. Die Veränderungsraten weisen also keinen signifikanten Trend über den Beobachtungszeitraum auf, und man könnte, um die Zahl der Parameter zu reduzieren, ein Modell mit der Restriktion $\mu = 0$ schätzen. Die Schätzergebnisse für die übrigen Parameter und die geglätteten Wahrscheinlichkeiten sind mit denen des Modells ohne diese Restriktion aber praktisch identisch, weswegen sie hier nicht dargestellt werden.

Anhang: Verwendete Testverfahren

1. LQ-Tests

Für das lineare Modell ist der maximale Wert der Log-Likelihood $L_0 = -169{,}36$ (vgl. Tabelle 1, S. 71). Daraus und aus den maximalen Werten L_1 für die Regimewechselmodelle, die in Kapitel 5 jeweils zusammen mit den geschätzten Parametern angegeben sind, lassen sich LQ-Teststatistiken folgendermaßen berechnen:

$$LQ = 2(L_1 - L_0) \ .$$

[175] In dieser Situation ist der Schwellenwert ja geschätzt.

2. Wald-Tests und asymptotische *t*-Tests

Wald-Teststatistiken lassen sich mit Hilfe der Schätzer der asymptotischen Varianz-Kovarianz-Matrix berechnen, die numerische Schätzverfahren wie der BHHH-Algorithmus am Konvergenzpunkt zur Verfügung stellen. Bei Verwendung des Befehls MAXIMIZE mit der Option BHHH unter *RATS for Windows* läßt sich dieser Schätzer als %XX ansprechen. Hier können Wald-Tests bequem mit den Befehlen TEST, RESTRICT und MRESTRICT durchgeführt werden.

Die zusammen mit den Schätzwerten jeweils angegebenen asymptotischen Standardabweichungen können für die Berechnung asymptotischer *t*-Tests von Restriktionen der Form $\theta = \theta_0$ eines skalaren Parameters θ verwendet werden. Die Wald-Teststatistik dieser Restriktion ist das Quadrat der entsprechenden *t*-Statistik.[176] Für große Stichproben ist die *t*-Statistik näherungsweise normalverteilt, also ihr Quadrat χ^2-verteilt. Dann sind asymptotischer *t*- und Wald-Test äquivalent. In kleinen Stichproben können aber aufgrund der Abweichung der *t*- von der Normalverteilung die Testergebnisse unterschiedlich ausfallen.

[176] Vgl. GREENE [1997], S. 164 f.

8 Prognosen

Ein wichtiger Vorzug der hier behandelten Regimewechselmodelle, bei denen die Gesetzmäßigkeiten der Parametervariation modelliert wird, gegenüber rein deskriptiven Verfahren ist die Möglichkeit der Erstellung von Prognosen, die die Natur der Parametervariation berücksichtigen. Wenn die entsprechenden Tests ergeben haben, daß ein Regimewechselmodell eine deutlich bessere Anpassungsgüte an die Daten besitzt als ein lineares Modell, so sollten sich damit auch Prognosen erstellen lassen, die denen eines linearen Modells überlegen sind. Die entsprechenden Prognoseformeln für Regimewechselmodelle werden in diesem Kapitel erläutert.

Das übliche Maß für die Qualität einer Prognose $\hat{y}_{T+r}(\mathbf{I}_T)$ einer Zufallsvariable y_{T+r} auf Basis der Realisation weiterer Zufallsvariablen $\mathbf{I}_T$ (der zum Zeitpunkt T verfügbaren Information) ist der mittlere quadratische Prognosefehler

$$E(\hat{y}_{T+r}(\mathbf{I}_T) - y_{T+r})^2 \ ,$$

gebildet mit Hilfe der gemeinsamen Verteilung von y_{T+r} und $\mathbf{I}_T$. Dieser mittlere quadratische Prognosefehler wird bekanntlich minimal, wenn y_{T+r} durch den bedingten Erwartungswert von y_{T+r} gegeben $\mathbf{I}_T$ prognostiziert wird, falls also

$$\hat{y}_{T+r}(\mathbf{I}_T) = E(y_{T+r} \mid \mathbf{I}_T) \ .$$

In der Praxis ist die gemeinsame Verteilung von y_{T+r} und $\mathbf{I}_T$ gewöhnlich unbekannt. Eine naheliegende und häufig verwendete Methode zur Auffindung von Prognosen besteht darin, anstelle der wahren Verteilung diejenige zu verwenden, die sich aus den geschätzten Parametern ergibt.

Häufig ist es nicht möglich, den bedingten Erwartungswert $E(y_{T+r} \mid \mathbf{I}_T)$ analytisch zu berechnen. In diesen Fällen kann man ihn durch Simulationen approximieren: Man erzeugt mit Hilfe eines Zufallsgenerators k unabhängige Realisationen der Störvariablen[177] und berechnet daraus unter Verwendung der Werte der beobachteten Variablen sowie der geschätzten Parameter die resultierenden Werte $\bar{y}_{T+r,1},\ldots,\bar{y}_{T+r,k}$ der zu prognostizierenden Variablen. Aus dieser simulierten bedingten Verteilung von y_{T+r} läßt sich der gesuchte bedingte Erwartungswert $E(y_{T+r} \mid \mathbf{I}_T)$ näherungsweise bestimmen als

$$E(y_{T+r} \mid \mathbf{I}_T) \approx \frac{1}{k}\sum_{i=1}^{k} \bar{y}_{T+r,i} \ .$$

[177] Üblich ist $k \geq 1000$.

Prognosefehler haben im allgemeinen verschiedene Ursachen:

- Fehlspezifikation. Dies ist wohl häufig die Hauptquelle von Prognosefehlern. Möglichst effiziente Testverfahren können helfen, ihren Einfluß zu vermindern.
- Abweichungen der geschätzten von den wahren Parametern. Diesen Fehler zu minimieren ist Aufgabe der Schätztheorie.
- Die Realisationen des Fehlerterms. Dieser Anteil läßt sich durch die Angabe von Prognoseintervallen um die Prognosewerte abschätzen. Diese Intervalle können in manchen Fällen berechnet werden, andernfalls kann man sie aus den oben beschriebenen Simulationen der bedingten Verteilung des zu prognostizierenden Wertes erhalten.

Die folgende Darstellung konzentriert sich auf die Prognose von n noch nicht realisierten bzw. nicht für die Schätzung herangezogenen Beobachtungen. Hier ist also[178]

$$\mathbf{I}_T = (y_1,\ldots,y_T,\mathbf{X}_1,\ldots,\mathbf{X}_{T+r},\mathbf{W}_1,\ldots,\mathbf{W}_{T+r})$$

bzw., falls der Schwellenindikator z_t exogen ist,

$$\mathbf{I}_T = (y_1,\ldots,y_T,\mathbf{X}_1,\ldots,\mathbf{X}_{T+r},\mathbf{W}_1,\ldots,\mathbf{W}_{T+r},z_1,\ldots,z_{T+r}) \ .$$

Nicht ausdrücklich betrachtet wird die Berechnung von *fitted values* (ex-post-Prognosen), d.h. der sich aus dem geschätzten Modell ergebenden Schätzwerte für die realisierten und für die Schätzung herangezogenen Beobachtungen. Für Schwellenmodelle könnte man sie ähnlich wie für lineare Modelle berechnen, da die Aufteilung der beobachteten Daten auf die Zustände (nach der Schätzung) bekannt ist. Für Mischungs- und Markov-Modelle ist ihre Berechnung in diesem Rahmen wenig sinnvoll, da bei der ML-Schätzung die Abfolge der Regime *nicht* geschätzt wird. Als *fitted values* kommen dann nur gewichtete Mittel der entsprechenden Schätzwerte aus den beiden Zuständen mit den geglätteten Zustandswahrscheinlichkeiten als Gewichte in Frage.

8.1 Schwellenmodelle

An dieser Stelle ist zu unterscheiden zwischen Prognosen, für die das Regime aus den Daten und dem (geschätzten) Schwellenwert ermittelt werden kann, und solchen, für die das nicht möglich ist.

Der erste Fall liegt vor, falls die Schwellenvariable exogen ist, oder wenn sie verzögert endogen vom Lag r' ist und $r \le r'$ Werte der endogenen Variablen prognostiziert werden sol-

[178] Die exogenen Größen werden natürlich für den Prognosezeitraum als bekannt vorausgesetzt. Falls das Modell verzögert endogene Regressoren der maximalen Laglänge p enthält, so enthalte $\mathbf{I}_T$ dementsprechend auch die Werte $y_{-p+1},\ldots,y_0$.

len. Ist der Schwellenwert c bekannt, so kann man auf den Zustand s_{T+r} zurückschließen, und es gilt

$$\begin{aligned} & E(y_{T+r} \mid \mathbf{I}_T) \\ &= E(y_{T+r} \mid s_{T+r}, \mathbf{X}_{T+r}, \mathbf{W}_{T+r}) \\ &= \begin{cases} \mathbf{W}_{T+r}\alpha + \mathbf{X}_{T+r}\beta_1, \text{falls } s_{T+r} = 1 \\ \mathbf{W}_{T+r}\alpha + \mathbf{X}_{T+r}\beta_2, \text{falls } s_{T+r} = 2 \end{cases} . \end{aligned}$$

Daraus erhält man Prognosen, indem man anstelle der wahren Parameter die geschätzten $\hat{\alpha}$, $\hat{\beta}_1$ und $\hat{\beta}_2$ verwendet. Bei unbekanntem Schwellenwert geht man genauso vor, nur daß man die Aufteilung der Daten auf die Zustände anhand des geschätzten Schwellenwertes $\hat{c}$ bestimmt.

Grundsätzlich anders stellt sich die Situation dar, falls der Regimeindikator verzögert endogen vom Lag r' ist und $r > r'$. In diesem Fall ist der Wert des Zustandsindikators, der das Regime für den zu prognostizierenden Wert bestimmt, nicht bekannt. Dann läßt sich die bedingte Verteilung dieses Wertes nicht mehr explizit berechnen.

Ein naives Prognoseverfahren besteht darin, den Zustand für den zu prognostizierenden Wert seinerseits aus dem prognostizierten Wert für den Schwellenindikator zu berechnen. Damit ignoriert man aber die stochastische Komponente des Modells vollkommen.[179] Ein kleiner Prognosefehler kann so zu einem falschen angenommenen Zustand für die folgenden Werte führen und damit zu unkalkulierbar hohen Fehlern in den weiteren Prognosen.

Besser ist es, die gesuchten bedingten Verteilungen wenigstens approximativ zu berechnen. Sie können entweder durch numerische Integration gewonnen werden,[180] oder man greift auf die oben beschriebenen Simulationsverfahren zurück. In CLEMENTS und SMITH [1997] wird ein Vergleich zwischen diesen und einigen weiteren Verfahren angestellt. Die Autoren kommen zu dem Schluß, daß die verhältnismäßig einfache, aber rechenintensive Simulationstechnik zu relativ guten Ergebnissen führt.

[179] Man berechnet nicht den Erwartungswert der durch die Strukturgleichung gegebenen Funktion der Störterme, sondern diese Funktion ihrer Erwartungswerte, die ja gleich Null sind. Es wird also nur das sogenannte „Skelett" des Modells, das aus der strukturellen Gleichung ohne Störeinfluß besteht, verwendet. Deswegen wird diese Methode auch „*SK method*" genannt, vgl. TONG [1995], S. 45, und CLEMENTS und SMITH [1997], S. 463.

[180] Vgl. hierzu z.B. TONG [1990], §4.2.4.

8.2 Mischungs- und Markov-Modelle

Falls die erklärenden Größen keine verzögert endogenen Variablen enthalten und der evt. vorhandene Regimeindikator z_t, von dem die Aufenthalts- bzw. Übergangswahrscheinlichkeiten abhängen, exogen ist, gilt

$$E(y_{T+r} \mid \mathbf{I}_T) = \sum_{s_{T+r}} E(y_{T+r} \mid s_{T+r}, \mathbf{X}_{T+r}, \mathbf{W}_{T+r}) P(s_{T+r} \mid \mathbf{I}_T),$$

wobei natürlich wieder

$$E(y_{T+r} \mid s_{T+r}, \mathbf{X}_{T+r}, \mathbf{W}_{T+r}) = \begin{cases} \mathbf{W}_{T+r}\alpha + \mathbf{X}_{T+r}\beta_1, \text{falls } s_{T+r} = 1 \\ \mathbf{W}_{T+r}\alpha + \mathbf{X}_{T+r}\beta_2, \text{falls } s_{T+r} = 2 \end{cases}.$$

Im Falle eines Mischungsmodells sind die Zustände unabhängig, also gilt im inhomogenen Fall

$$P(s_{T+r} \mid \mathbf{I}_T) = P(s_{T+r} \mid z_{T+r}) = \begin{cases} f(\gamma\, z_{T+r} + \delta), \text{falls } s_{T+r} = 1 \\ 1 - f(\gamma\, z_{T+r} + \delta), \text{falls } s_{T+r} = 2 \end{cases}$$

und im homogenen Fall

$$P(s_{T+r} \mid \mathbf{I}_T) = P(s_{T+r}) = \begin{cases} p, \text{falls } s_{T+r} = 1 \\ 1 - p, \text{ falls } s_{T+r} = 2. \end{cases}$$

Die Prognose von y_{T+r} ist also das Mittel aus den beiden Prognosen, die man bei Kenntnis des Zustandes s_{T+r} verwenden würde, gewichtet mit der aus den geschätzten Parametern berechneten Wahrscheinlichkeit für diesen Zustand.

Im Falle eines Markov-Modells muß man diese Wahrscheinlichkeit mit Hilfe der Dynamik des Markov-Kette auf den letzten Zeitpunkt zurückführen, für den man Beobachtungen über die abhängige Variable hat:

$$P(s_{T+r} \mid \mathbf{I}_T) =$$
$$\sum_{s_{T+r-1}} \cdots \sum_{s_T} P(s_{T+r} \mid s_{T+r-1}, z_{T+r}) \cdots P(s_{T+1} \mid s_T, z_{T+1})$$
$$P(s_T \mid y_1, \ldots, y_T, \mathbf{W}_1, \ldots, \mathbf{W}_T, \mathbf{X}_1, \ldots, \mathbf{X}_T).$$

Die gesuchte Wahrscheinlichkeit erhält man demzufolge aus den geschätzten Übergangswahrscheinlichkeiten und den geglätteten Wahrscheinlichkeiten zum Zeitpunkt T, wie sie z.B. im letzten Schritt des EM-Algorithmus berechnet werden.

Komplizierter sind die Prognoseformeln, falls unter den erklärenden Größen verzögert endogene Variablen sind. Man muß dann die eben erläuterte Methode kombinieren mit rekursiven Verfahren zur Berechnung von Prognosen für lineare AR-Prozesse.[181]

[181] Genaueres hierzu findet sich in KROLZIG [1997], S. 65 ff.

Ist schließlich der Regimeindikator verzögert endogen, so ist die Situation ähnlich wie für Zeitreihen-Schwellenmodelle (SETAR-Modelle): Solange der Wert des Indikators für den zu prognostizierenden Wert bekannt ist, behalten die oben beschriebenen Formeln ihre Gültigkeit. Übersteigt der Prognosehorizont r jedoch die Laglänge des Zustandsindikators, läßt sich die gesuchte bedingte Verteilung nicht mehr explizit berechnen, und man ist auf numerische Integrationen oder Simulationen angewiesen.

9 Prognosen für Wochendaten des FF/DM-Wechselkurses

Um die in Kapitel 8 beschriebenen Prognoseverfahren an dem in Kapitel 7 ausgewählten Modell für den FF/DM-Wechselkurs, dem Markov-Modell, zu erläutern, werden zwei Prognosen über einen Zeitraum von je 20 Wochen dargestellt, beginnend zu einem Zeitpunkt mit hoher bzw. niedriger Volatilität. Die letzte ausgedehnte Phase mit hoher Volatilität war Ende 1995. Als Beginn der ersten Prognose wurde der 15. September 1995 gewählt. Die zweite Prognose beginnt an der 21. Beobachtung vor Ende des Beobachtungszeitraums, also am 26. Juni 1998, so daß noch 20 Werte für den Vergleich der Prognosen mit den realisierten Werten zur Verfügung standen. Das Markov-Modell wurde für die entsprechenden Zeiträume neu geschätzt. Die Schätzergebnisse, die sich nicht wesentlich von denen in Tabelle 5 (S. 79) unterscheiden, sind in Tabelle 13 wiedergegeben.[182]

Prognose	$\hat{p}^{11}$	$\hat{p}^{22}$	$\hat{\mu}$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
1	0,832 (0,057)	0,970 (0,009)	-0,0063 (0,0102)	0,767 (0,070)	0,178 (0,007)
2	0,835 (0,060)	0,954 (0,015)	-0,0071 (0,0075)	0,634 (0,094)	0,150 (0,014)

Tabelle 13: Geschätzte Parameter eines Markov-Modells mit Schätzzeitraum 23. Januar 1987 bis 15. September 1995 / 26. Juni 1998

Für die Prognose sind zuerst nach der auf S. 120 angegebenen Formel aus der geglätteten Wahrscheinlichkeit am letzten Beobachtungszeitpunkt und den geschätzten Übergangswahrscheinlichkeiten die Wahrscheinlichkeiten für die beiden Zustände an den Prognosezeitpunkten zu ermitteln. Daraus lassen sich dann Prognosen für den Wert von r_t und Prognoseintervalle berechnen.

Prognose 1

Für die erste Prognose haben die Wahrscheinlichkeiten für den Zustand mit der hohen Volatilität den in Abbildung 12 dargestellten Verlauf. Die prognostizierte Wahrscheinlichkeit nähert sich, wie es die Theorie vorhersagt,[183] exponentiell der unbedingten Wahrscheinlichkeit für diesen Zustand, die in dieser Schätzung 0,151 beträgt. Auf sehr lange Sicht lautet die Prognose also, daß das Modell im stationären Zustand ist, und die Prognosen werden zeitunabhängig. Für kürzere Prognosehorizonte macht sich die hohe geglättete Wahrscheinlichkeit am letzten Zeitpunkt des Schätzzeitraumes in höheren prognostizierten Wahrscheinlichkeiten bemerkbar.

[182] Die Testergebnisse für diese Regimewechselmodelle führten zu denselben Ergebnissen wie in Kapitel 7.

[183] Vgl. den Anhang von Kapitel 2, besonders S. 20.

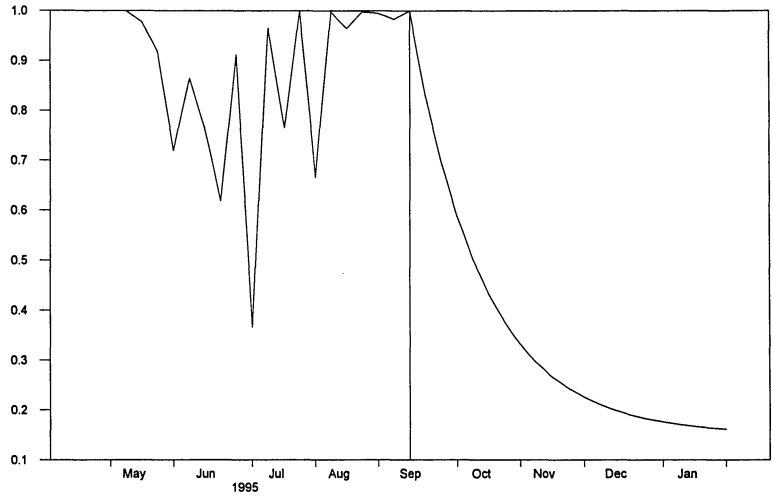

Abbildung 12: Geglättete und prognostizierte Wahrscheinlichkeiten für den Zustand mit der hohen Volatilität (Prognose 1)

Die Prognose des Wertes der Variablen ist in diesem Fall nur von geringerem Interesse, da sie gleich dem mit den prognostizierten Wahrscheinlichkeiten gewichteten Mittel der Mittelwerte in den beiden Zuständen ist. Diese Mittelwerte sind aber gleich, weswegen der prognostizierte Wert gar nicht von den geschätzten Wahrscheinlichkeiten abhängig ist, also zeitlich konstant ist. Außerdem unterscheidet er sich nur geringfügig von der Prognose des linearen Modells (vgl. Abbildung 13).

Von größerer Bedeutung in diesem Modell, das sich ja wesentlich in den geschätzten Standardabweichungen in den beiden Zuständen unterscheidet, ist der Verlauf des Prognoseintervalls. Will man ein solches Prognoseintervall für den prognostizierten Wert berechnen, so benötigt man die geschätzte Verteilung der zu prognostizierenden Werte. Dies ist in diesem Fall eine Mischverteilung, deren Gewichte für einen Prognosezeitpunkt gerade die prognostizierten Wahrscheinlichkeiten darstellen. Für sie lassen sich numerisch Prozentpunkte und damit Prognoseintervalle bestimmen. In Abbildung 13 sind diese Intervalle für die Prognosewahrscheinlichkeit 0,95 den beobachteten Werten sowie den Prognoseintervallen des linearen Modells gegenübergestellt.

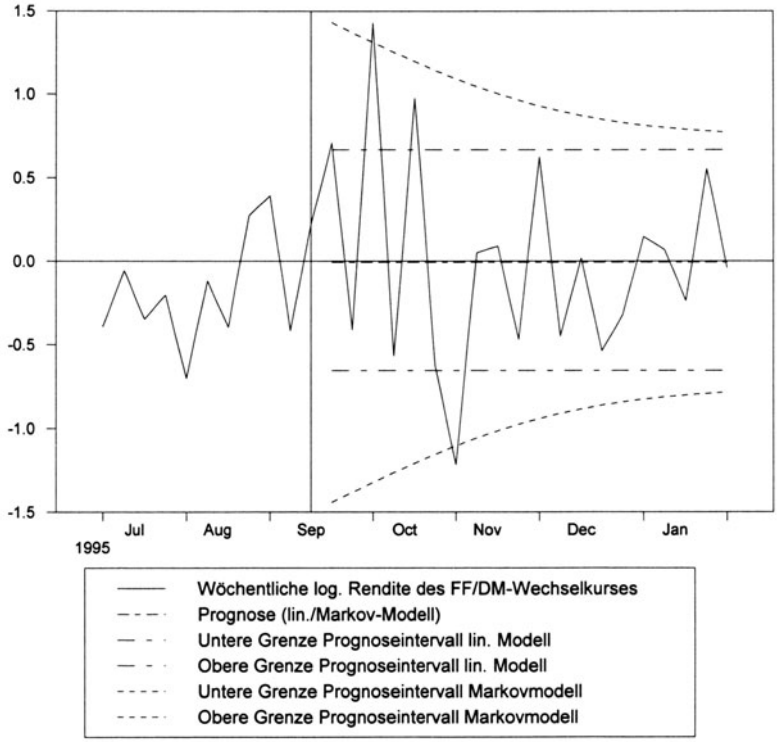

Abbildung 13: 95 %-Prognoseintervalle des linearen und des Markov-Modells (Prognose 1)
(Die Prognosen der Werte von r_t mit den beiden Modellen sind in der Darstellung nicht unterscheidbar, da zu ähnlich.)

Man sieht, daß das Prognoseintervall des Markov-Modells für die ersten Prognosezeitpunkte wesentlich breiter als das des linearen Modells ist. Der Grund ist die sich im Markov-Modell widerspiegelnde Persistenz des Zustandes mit der hohen Volatilität. Parallel zur Abnahme der prognostizierten Wahrscheinlichkeit für diesen Zustand wird auch das Prognoseintervall kleiner. Tatsächlich folgten auf den Beginn der Prognose noch einige sehr hohe Ausschläge des Wechselkurses, von denen zwei durch keines der Prognoseintervalle und zwei durch das des Markov-Modells, nicht aber durch das des linearen Modells, erfaßt werden.

Dies ist natürlich ein isoliertes Ergebnis, das keineswegs repräsentativ für die Prognosequalität der Modelle sein muß. Es zeigt aber prinzipiell, wie das Markov-Modell in der Lage ist, die Breite von Prognoseintervallen an die Volatilität an den letzten Beobachtungspunkten anzupassen.

Prognose 2

Für die zweite Prognose ist die Situation umgekehrt: Am Ende des Schätzzeitraums für diese Prognose, also am 26. Juni 1998, ist die geglättete Wahrscheinlichkeit für den Zustand mit der hohen Volatilität sehr niedrig (etwa 0,003), da schon über einen längeren Zeitraum keine größeren Wechselkursausschläge mehr aufgetreten sind. Die prognostizierten Wahrscheinlichkeiten für diesen Zustand sind zusammen mit den geglätteten Wahrscheinlichkeiten bis zum Beginn der Prognose in Abbildung 14 dargestellt. Sie wachsen exponentiell an bis zur unbedingten Wahrscheinlichkeit dieses Zustandes, die mit den geschätzten Parametern bis zu diesem Zeitpunkt 0,216 beträgt.

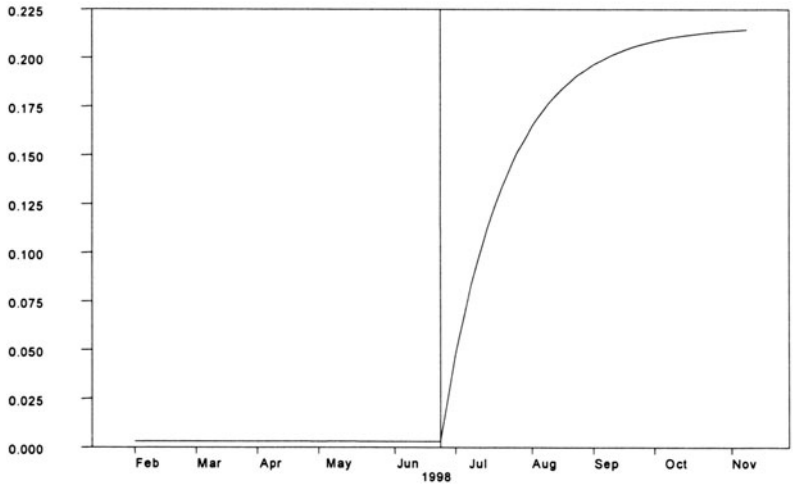

Abbildung 14: Geglättete und prognostizierte Wahrscheinlichkeiten für den Zustand mit der hohen Volatilität (Prognose 2).
(Die geglätteten Wahrscheinlichkeiten schwanken in der ersten Hälfte 1998 so wenig, daß sie hier fast konstant erscheinen.)

Dementsprechend ist das Prognoseintervall, das in Abbildung 15 dargestellt ist, erst kleiner als das des linearen Modells und wächst dann an. Deutlich zu sehen ist hier, daß das Prognoseintervall für Prognosehorizonte von mehr als einem Monat *breiter* ist als das des linearen Modells.

Der Grund ist, daß die prognostizierten Wahrscheinlichkeiten für die Zustände gegen die unbedingten Wahrscheinlichkeiten konvergieren (vgl. Abbildung 14). Prognoseintervalle für lange Prognosezeiträume entsprechen also näherungsweise denen von Werten im stationären Zustand. Die prognostizierte Verteilung eines solchen Zustandes ist eine Mischungsvertei-

lung, deren Gewichte die unbedingten Wahrscheinlichkeiten der Zustände sind. Diese Verteilung ist, wie in Kapitel 5.3 erläutert wurde, sehr ähnlich zu der des Mischungsmodells aus Kapitel 5.2, die mehr Wahrscheinlichkeitsmasse auf den *tails* als die Normalverteilung hat, die die Basis für die Prognosen des linearen Modells darstellt (vgl. Abbildung 8, S. 77). Um dieselbe Prognosewahrscheinlichkeit zu erhalten, muß das Prognoseintervall des Markov-Modells breiter als das des linearen Modells sein. Der Grund dafür, daß die Prognoseintervalle größer werden als die des linearen Modells, ist also, daß das im Vergleich zur Normalverteilung zu häufige Auftreten von hohen absoluten Werten vom Markov-Modell modelliert wird.

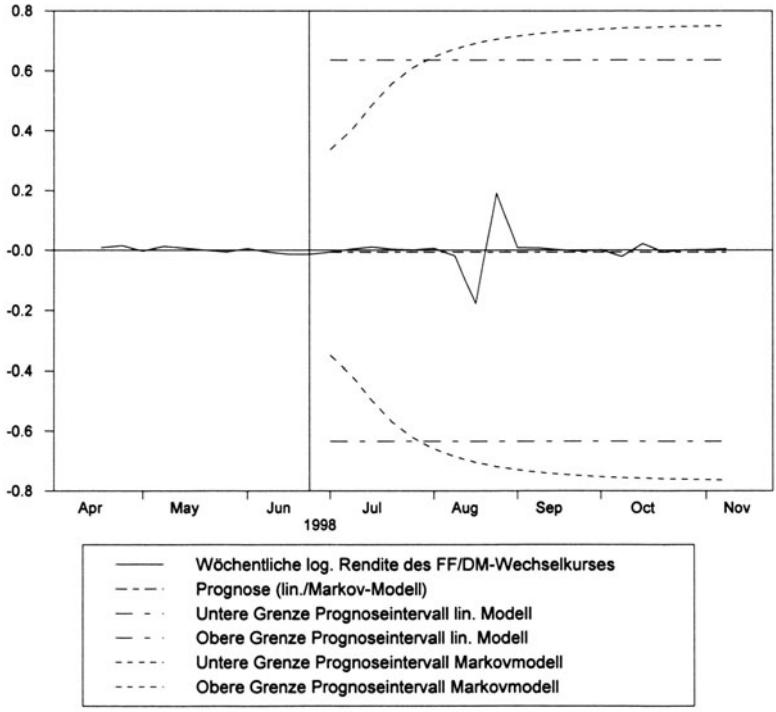

Abbildung 15: 95 %-Prognoseintervalle des linearen und des Markov-Modells (Prognose 2)
(Die Prognosen der Werte von r_t der beiden Modelle sind in der Darstellung nicht unterscheidbar, da zu ähnlich.)

10 Weitere empirische Ergebnisse und ökonomische Implikationen

10.1 Untersuchung der Wochendaten

10.1.1 Zusammenfassung der Ergebnisse für den FF/DM-Kurs und ökonomische Schlußfolgerungen

Die Analyse in Kapitel 5 und 7 hat gezeigt, daß sich die Leptokurtosis der Verteilung der FF/DM-Renditen und die Autokorrelation der quadrierten Beobachtungen mit einem heteroskedastischen Markov-Modell beschreiben lassen. Für die Prognose der *Werte* stellt dieses Modell keinen Gewinn dar, wohl aber für die Vorhersage von Prognoseintervallen, d.h. die Abschätzung der künftigen Volatilität.

Daraus ergeben sich wichtige Konsequenzen für die Finanzmarkttheorie, in der in der Regel von unabhängig normalverteilten Renditen ausgegangen wird, wie etwa im *Capital Asset Pricing Model* (CAPM). Die Gültigkeit des Markov-Modells hat auch Implikationen für die Bewertung von in DM notierten Optionen auf Französische Francs (und andersherum). Die Standardformel zur Bewertung von Kaufoptionen, die Black-Scholes-Formel, ist gültig, wenn die Renditen unabhängig identisch normalverteilt sind. Bei Vorliegen einer Mischverteilung muß diese Formel modifiziert werden, wobei dieser Fall eine relativ einfache analytische Lösung besitzt.[184] Diese Optionspreisformel reproduziert den aus empirischen Untersuchungen bekannten "Smile-Effekt", also die systematische Abweichung der beobachteten Optionspreise von den Black-Scholes-Werten für kurze Restlaufzeiten und einem Verhältnis von Aktien- zu Basispreis nahe bei Eins. Für Markov-Modelle berechnen KAEHLER und MARNET [1994] Optionspreise mit Hilfe von Simulationen und kommen zu ähnlichen Ergebnissen.

10.1.2 Der NG/DM-Kurs

Im folgenden soll der Verlauf eines weiteren Wechselkurses im EWS anhand von Wochendaten untersucht werden, der zwischen dem Niederländischen Gulden (NG) und der Deutschen Mark. Zwischen diesen Währungen fanden während der Dauer des EWS nur insgesamt 3 Realignments statt, davon 2 zu Beginn des EWS 1979 und das letzte 1983. Seitdem ist der NG/DM-Leitkurs nicht mehr verändert worden und schließlich in die NG/DM-Relation im Rahmen der Europäischen Währungsunion übergegangen.

[184] Vgl. RITCHEY [1990].

Abbildung 16: Wechselkurs zwischen Gulden und DM, Leitkurs und Interventionsgrenzen

Der Verlauf des Wechselkurses, der für denselben Zeitraum wie der zwischen Franc und DM (23. Januar 1987 bis 20. November 1998) in Abbildung 16 dargestellt ist,[185] war während der gesamten Dauer des EWS wesentlich stabiler als der zwischen Franc und DM. Nie gelangte der Kurs ernsthaft in die Nähe einer der Interventionsgrenzen, so daß das Auftreten von Spekulation gegen eine der beteiligten Zentralbanken zu keinem Zeitpunkt wahrscheinlich war. Daraus läßt sich die Hypothese ableiten, daß ein lineares Modell in diesem Fall die Daten besser beschreibt als für den FF/DM-Wechselkurs. Für die folgende Untersuchung wurden wieder die logarithmischen Differenzen von wöchentlichen Durchschnittswerten zugrunde gelegt. In Tabelle 14 sind wichtige statistische Daten der Verteilung dieser Differenzen wiedergegeben.

[185] Am 2. August 1993 wurden die obere Interventionsgrenze auf 130,834 NG/DM herauf- und die untere Interventionsgrenze auf 97,0325 NG/DM herabgesetzt. Diese Kurse fanden aber keine Anwendung. Aufgrund einer bilateralen Vereinbarung zwischen den Niederlanden und der Bundesrepublik Deutschland wurden die alten, engeren Bandbreiten aufrechterhalten, die deshalb auch in der Grafik dargestellt sind. Quelle: Datastream.

	Wert	Signifikanzniveau
Minimum	-0,117	
Maximum	0,139	
Maximaler Wert der Log-Likelihood	1330,87	
Mittelwert μ	0,000033	
Standardabweichung σ	0,0281	
Schiefe	0,262	0,008
Kurtosis	2,82	0,00
Jarque-Bera	211,57	0,00
Durbin-Watson	2,45	<0,01
LB-10 / 20 (für r_t)[a]	47,95 / 66,79	0,00 / 0,00[b]
Dickey-Fuller	-11,66	<0,01
LB-10 / 20 (für r_t^2)[a]	60,19 / 80,24	0,00 / 0,00[c]
ARCH(1)	140,28	0,00

Tabelle 14: Statistische Daten der Verteilung von r_t
[a]: LB-10 / 20: Ljung-Box-Test bis zum 10. / 20. Lag
[b]: Zum 5 %-Niveau signifikant: 1., 4. und 19. Lag
[c]: Zum 5 %-Niveau signifikant: 1. bis 3. und 12. Lag

Ein Vergleich der Standardabweichung und der Spannweite der Daten mit den entsprechenden Werten in Tabelle 1 (S. 71) zeigt, daß die Ausschläge des NG/DM-Wechselkurses deutlich niedriger waren als die des FF/DM-Kurses. Die Schiefe der Werte ist etwas höher, die Kurtosis jedoch niedriger als die der FF/DM-Daten. Beide sind jedoch signifikant von Null verschieden, und der Jarque-Bera-Test verwirft die Nullhypothese einer Normalverteilung der Daten. Interessant ist, daß der Durbin-Watson-Test und der Ljung-Box-Test die Nullhypothese der Unkorreliertheit verwerfen. Die empirische Autokorrelation erster Ordnung ist mit einem t-Wert von –5,61 (dem absolut gesehen höchsten der t-Werte aller geschätzten Autokorrelationen) hochsignifikant. Der Dickey-Fuller-Test verwirft die Nullhypothese eines Random Walks. Der Ljung-Box-Test für die quadrierten Beobachtungen und der ARCH(1)-Test schließlich zeigen signifikante Autokorrelation der quadrierten Werte an. Im Vergleich zu den FF/DM-Daten kommt hier also zu den *stylised facts* ein weiteres hinzu, nämlich die Autokorrelation der Beobachtungen.

Schätzt man ein AR(1)-Modell der Form

$$r_t = \varphi\, r_{t-1} + u_t \ ,$$

so erhält man die in Tabelle 15 dargestellten Schätzergebnisse. Die Werte der Durbin-Watson- und der Ljung-Box-Teststatistik (für die Residuen) zeigen, daß *ein* Lag in dem autoregressiven Modell genügt, um die Autokorrelation in den Daten ausreichend zu modellieren.

Charakteristisch ist der negative Autokorrelationskoeffizient 1. Ordnung, der bedeutet, daß auf positive Ausschläge mit einer gewissen Regelmäßigkeit negative Ausschläge folgten.

$\hat{\varphi}$	$\hat{\sigma}$	DW	LB-10 / 20	Maximaler Wert der Log-Likelihood
-0,226 (0,039)	0,0274	1,98 (>0,05)[a]	9,5 / 29,7 (0,48 / 0,08)[a]	1344,81

Tabelle 15: Geschätzte Parameter eines AR(1)-Modells für r_t
[a]: Signifikanzniveau

Nimmt man dieses lineare Modell als Ausgangsbasis, so kann man versuchen, die Leptokurtosis und die bedingte Heteroskedastizität mit AR(1)-Regimewechselmodellen zu erklären.

Der maximale Wert der Log-Likelihood des entsprechenden Schwellenmodells mit dem Zins als Schwellenindikator ist 1350,02. Der Wert der LQ-Statistik ist damit 10,42. Auf Basis der Faustregel aus Kapitel 6.2.2.1.2 muß man diesen Wert mit den kritischen Werten einer $\chi^2(5)$-Verteilung vergleichen. Damit liegt der p-Wert zwischen 0,05 und 0,10. Die Nullhypothese ist also in diesem Fall nicht zu verwerfen.

Anders ist die Situation für Mischungs- und Markov-Modelle. Die Werte der LQ-Statistik der Tests der Nullhypothese des AR(1)-Modells gegen die Alternative eines AR(1)-Mischungs- bzw. eines AR(1)-Markov-Modells lauten 85,68 bzw. 140,34. Da für diese Modelle die exakte Verteilung der Teststatistiken nicht bekannt ist und keine simulierten Werte vorliegen, ist man auf die Faustregel aus Kapitel 6.2.2.2 angewiesen. Die kritischen Werte einer $\chi^2(5)$-Verteilung und einer $\chi^2(6)$-Verteilung zum 1 %-Niveau lauten 15,09 und 16,81. Man kann also auch ohne genaue Kenntnis der Verteilung der Teststatistik unter der Nullhypothese mit hinreichender Sicherheit davon ausgehen, daß die Nullhypothese in beiden Fällen verworfen werden muß.

Der Wert der LQ-Statistik für den Test der Nullhypothese des Mischungs- gegen die Alternative des Markov-Modells lautet 54,66. Hier sind die kritischen Werte einer $\chi^2(1)$-Verteilung zu verwenden. Die Nullhypothese wird also zu allen üblichen Niveaus verworfen. Die Schätzwerte des aus der Testkaskade hervorgehenden AR(1)-Markov-Modells sind in Tabelle 16 dargestellt.

$\hat{p}^{11}$	$\hat{p}^{22}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,863 (0,044)	0,894 (0,034)	-0,282 (0,057)	-0,044 (0,059)	0,038 (0,004)	0,014 (0,002)	1414,98

Tabelle 16: Schätzwerte eines AR(1)-Markov-Modells für den NG/DM-Kurs

Auch für den (im Vergleich zum FF/DM-Kurs sehr stabilen) NG/DM-Kurs lassen sich also verschiedene Zustände ausmachen. Diese unterscheiden sich zum einen um einen Faktor von etwas mehr als 2 in den Standardabweichungen, also etwas weniger stark als für den FF/DM-Kurs. Der Wert der Wald-Teststatistik der Nullhypothese eines homoskedastischen Modells ist 170,65. Die Nullhypothese wird somit verworfen. Zum anderen sind die geschätzten autoregressiven Parameter in den beiden Zuständen deutlich verschieden. Der Wert der Wald-Teststatistik der Nullhypothese eines Modells mit zustandsunabhängigen autoregressiven Parametern ist 5,32 mit einem p-Wert von 0,02. Damit wird die Nullhypothese zum Niveau 5 % verworfen.

Der geschätzte Fehler von $\hat{\varphi}_2$, also dem autoregressiven Parameter in dem Zustand mit der niedrigen Volatilität, ist relativ groß im Vergleich zu dem Schätzwert des Parameters. Der t-Wert ist –0,74 mit einem p-Wert von 0,46. Es bietet sich also an, das Modell mit der Restriktion $\varphi_2 = 0$ zu schätzten.[186] Die Schätzergebnisse sind in Tabelle 17 wiedergegeben.

Die Schätzergebnisse der übrigen Parameter sind kaum verändert gegenüber dem unrestringierten Markov-Modell, ihre geschätzten Standardfehler sind durchgängig kleiner. Der Wert der LQ-Statistik für die Restriktion $\varphi_2 = 0$ ist 0,54, der LQ-Test führt damit zu demselben Ergebnis wie der zuletzt erwähnte t-Test.

$\hat{p}^{11}$	$\hat{p}^{22}$	$\hat{\varphi}_1$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,865 (0,030)	0,891 (0,022)	-0,287 (0,055)	0,038 (0,001)	0,014 (0,001)	1414,71

Tabelle 17: Schätzwerte eines AR(1)-Markov-Modells für den NG/DM-Kurs mit der Restriktion $\varphi_2 = 0$

Der geschätzte autoregressive Parameter $\hat{\varphi}_1$ ist noch stärker negativ als der des linearen Modells (-0,226, vgl. Tabelle 15). Der AR(1)-Mechanismus scheint also tatsächlich nur in Zustand 1 zu wirken, dort aber stärker, als es das lineare Modell, in dem die Parameter aus Beobachtungen aus beiden Zuständen geschätzt (also gemittelt) werden, angibt. Dies ist der Zustand mit der hohen Volatilität. Offensichtlich folgten *großen* Kursausschlägen regelmäßig Kursveränderungen in die jeweils andere Richtung, nicht aber kleinen Ausschlägen. Dies ist ein Ausdruck der starken Bindung des Gulden an die DM: Auf größere Kursbewegungen in Richtung der Interventionsgrenzen reagierten Marktteilnehmer (darunter u.U. auch die Zentralbank der Niederlande[187]) sofort mit Käufen bzw. Verkäufen, die den Kurs rasch wieder in

[186] Hier liegt also der Fall vor, in dem ein Regressor nur in einem der beiden Zustände Einfluß auf die abhängige Variable ausübt.

[187] Die Bundesbank beteiligte sich ja prinzipiell nicht an marginalen Interventionen, vgl. S. 23.

die Nähe des Leitkurses brachten. Nach größeren Ausschlägen war die Richtung der Kursentwicklung also kurzfristig in einem gewissen Maße vorhersagbar. Kleinere Kursausschläge wurden aber von den Marktteilnehmern toleriert. Das Regimewechselmodell läßt die negative Autokorrelation der Daten in einem neuen Licht erscheinen und enthüllt eines der in Kapitel 3.3 erwähnten wiederkehrenden Muster des Kursverlaufs.

Die Bedingung $r_t = 0$ in Zustand 2 ließe sich (wie für den FF/DM-Kurs) durch die Restriktion $r_t = \mu_1$ ersetzen.[188] Schätzt man ein Modell mit diesem bedingten Mittelwert in Zustand 2, bei ansonsten aber unveränderter Parametrisierung, so erhält man praktisch dieselben Parameterschätzer wie in Tabelle 17 und zusätzlich $\hat{\mu}_1 = -0{,}000067$ mit einem t-Wert von $-0{,}075$. Der Drift in Zustand 2 ist also nicht signifikant. Die geglätteten Wahrscheinlichkeiten für Zustand 1 des Modells aus Tabelle 17 sind in Abbildung 17 dargestellt.[189]

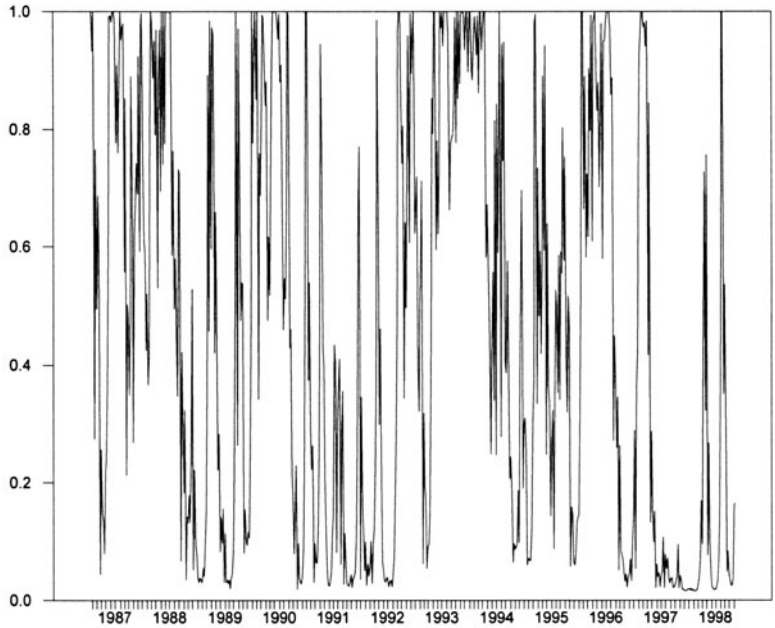

Abbildung 17: Geglättete Wahrscheinlichkeit für Zustand 1 (Markov-Modell mit AR(1)-Mechanismus in Zustand 1)

[188] Dies wäre der Fall, in dem in den beiden Zuständen unterschiedliche Regressoren die endogene Variable erklären, nämlich in dem einen eine Konstante und in dem anderen die verzögert endogene Variable.

[189] Sie werden in Kapitel 10.1.3 genauer untersucht.

Die Implikationen für die Finanzmarkttheorie, die in Kapitel 10.1.1 erwähnt wurden, gelten sinngemäß auch hier, wobei aber noch die negative Autokorrelation in einem der beiden Zustände zu berücksichtigen ist. Sie hat zur Konsequenz, daß hier auch die *Werte* in einem gewissen Maße prognostizierbar waren (nämlich nach hohen Kursausschlägen). Daraus ließ sich leicht eine profitable Handelsstrategie für die Marktteilnehmer gewinnen, die umgekehrt den Marktteilnehmern Verluste beschert haben wird, die in diesem Sinne vorhersagbar handelten. Dies dürfte vorwiegend die Zentralbank der Niederlande gewesen sein, die ein Interesse daran hatte, den Wechselkurs in der Nähe des Leitkurses zu halten.[190] Es ist also zu vermuten, daß der Zentralbank der Niederlande aus der Kopplung der beiden Währungen Kosten entstanden sind.

10.1.3 Vergleich der Modelle für die beiden Währungen.

Der Zustand mit der hohen Volatilität muß in den beiden Modellen sicherlich unterschiedlich interpretiert werden. Der Franc näherte sich mehrfach der oberen Interventionsgrenze (vgl. Abbildung 4, S. 69), so daß ein Realignment durchaus im Bereich des Möglichen lag. Hier ist also davon auszugehen, daß – zumindest bis zur Erweiterung der Bandbreiten am 2. August 1993 – tatsächlich Spekulation auf eine FF-Abwertung eine Rolle gespielt haben dürfte. Dies gilt so sicherlich nicht für den Gulden, der nie ernsthaft in die Nähe einer der Interventionsgrenzen gelangt ist.

Dennoch zeigt ein Vergleich der geglätteten Wahrscheinlichkeiten für den Zustand mit der hohen Volatilität für den FF/DM- und den NG/DM-Kurs (vgl. Abbildung 10, S. 80, und Abbildung 17, S. 132) einen auffälligen Gleichlauf der Entwicklung. Diese Wahrscheinlichkeit war in beiden Fällen hoch von Ende 1987 bis Anfang 1988, von Ende 1989 bis Ende 1990, von Mitte 1992 bis Mitte 1994 und von Anfang 1995 bis Ende 1996. Die "Unruhe" auf den Märkten übertrug sich also durchaus von einer abwertungsgefährdeten Währung auf eine als stabil angesehene (oder andersherum).

10.2 Eine Empirische Überprüfung der relativen Kaufkraftparitätentheorie am Beispiel von Monatsdaten des FF/DM- und des NG/DM-Kurses

Die Mittelwerte der Veränderungsraten des FF/DM- und des NG/DM-Kurses wurden in den bisher dargestellten Untersuchungen nicht bzw. durch die verzögert endogene Variable er-

[190] Die Bundesbank beteiligte sich ja nicht an intramarginalen Interventionen, vgl. 3.1.2.2.

klärt. Dies ist vom theoretischen Standpunkt unbefriedigend: Die Kaufkraftparitätentheorie (vgl. Kapitel 3.2.1.1) in ihrer relativen Form besagt, daß diese Rate gleich der Differenz der Veränderungsraten der Güterpreise ist.

Die vorherrschende Meinung ist, daß die PPP – wenn überhaupt – nur für langfristige, jedoch nicht für kurzfristige Veränderungsraten empirisch nachgewiesen werden kann. Dies soll in einer abschließenden Untersuchung anhand des FF/DM- und des NG/DM-Kurses im Kontext von Regimewechselmodellen überprüft werden.

Entsprechende Preisindizes sind (zum gegenwärtigen Zeitpunkt) nur in Monatswerten bis einschließlich Januar 1998 verfügbar. Grundlage der Untersuchung bilden also die mit 100 multiplizierten monatlichen logarithmischen Veränderungsraten der Wechselkurse und der Verbraucherpreisindizes (CPI) für die jeweiligen Länder im Zeitraum von Februar 1987 bis Januar 1998 (132 Beobachtungen).[191]

Sei $r_t = 100 \cdot \log(w_t / w_{t-1})$ und $x_t = 100 \cdot (\log(P_t^* / P_{t-1}^*) - \log(P_t / P_{t-1}))$. Ausgangspunkt der Untersuchungen bildet die Gleichung

$$r_t = \alpha + \beta x_t + u_t$$

mit $\mathbf{u} \sim N(0, \sigma^2 \mathbf{I})$. Die Gültigkeit der PPP erfordert $\alpha = 0$ und $\beta = 1$.

Da es sich um differenzierte Daten handelt, ist es nicht verwunderlich, daß der Dickey-Fuller-Test mit einem Wert von -7,05 für r_t und von -4,29 für x_t die Nullhypothese der Integration in beiden Fällen verwirft. Schätzt man die Gleichung nach der OLS-Methode, so erhält man in beiden Fällen nicht signifikante Konstanten. In Tabelle 18 sind deshalb die Schätzergebnisse mit der Restriktion $\alpha = 0$ für die beiden untersuchten Wechselkurse dargestellt.

	$\hat{\beta}$	$\hat{\sigma}$	DW	R^2	Maximaler Wert der Log-Likelihood
FF/DM	0,14 (0,17)	0,267	2,05	0,005	-12,25
NG/DM	0,018 (0,020)	0,047	2,29	0,006	214,57

Tabelle 18: OLS-Schätzergebnisse der relativen PPP (ohne Konstante)

Die Ergebnisse widersprechen in beiden Fällen klar der PPP. Die geschätzten Koeffizienten sind zwar positiv, aber zu klein und nicht signifikant. Die Schätzanpassung ist indiskutabel.

Die bisherigen Untersuchungen haben gezeigt, daß bei beiden Wechselkursen offensichtlich sprunghafte Verhaltensänderungen einen Teil der Charakteristika der Daten erklären

[191] Quelle: Datastream, IMF.

können. Deswegen wurde die PPP-Gleichung mit verschiedenen Regimewechselmodellen neu geschätzt.

Für den NG/DM-Kurs war keines der untersuchten Regimewechselmodelle signifikant. Daraus muß geschlossen werden, daß die enge Bindung des Gulden an die DM und die damit verbundenen Zentralbankeingriffe in diesem Fall keinen Zusammenhang zwischen Wechselkurs- und Preisentwicklung zugelassen haben.

Für den FF/DM-Kurs waren verschiedene Regimewechselmodelle signifikant, mit denen dasselbe Testverfahren durchgeführt wurde wie mit den Wochendaten. Es führte zu einem Markov-Modell, dessen Regressionsparameter β_2 in Zustand 2 gleich Null ist. Die übrigen geschätzten Koeffizienten sind in Tabelle 19 angegeben.

$\hat{p}^{11}$	$\hat{p}^{22}$	$\hat{\beta}_1$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	Maximaler Wert der Log-Likelihood
0,60 (0,31)	0,91 (0,07)	1,08 (0,35)	0,50 (0,07)	0,16 (0,02)	6,71

Tabelle 19: Schätzergebnisse des Markov-Modells (PPP) für den FF/DM-Kurs

Die Wahrscheinlichkeit für das Verbleiben in Zustand 2 ist höher als für Zustand 1. Dementsprechend ist die unbedingte (stationäre) Wahrscheinlichkeit für Zustand 1 relativ niedrig (0,18), und die erwartete Verweildauer in Zustand 1 mit 2,5 Monaten geringer als in Zustand 2 mit 11,2 Monaten. Interessant ist der geschätzte Koeffizient $\hat{\beta}_1$ in Zustand 1: Er ist signifikant von Null verschieden und liegt überraschend nahe an dem theoretisch vorhergesagten Wert von 1. Ein Wald-Test (d.h. ein asymptotischer t-Test) verwirft die Nullhypothese $\beta_1 = 1$ nicht.

Die Ergebnisse lassen die Interpretation zu, daß die PPP in gewissen Phasen tatsächlich einen Erklärungsgehalt für die Wechselkursveränderungen besessen hat. Diese Phasen machten aber nur einen verhältnismäßig geringen Anteil der Beobachtungen aus (gemessen an der unbedingten Wahrscheinlichkeit für Zustand 1 etwa 1/5 der Werte). Die geglätteten Wahrscheinlichkeiten für Zustand 1, die in Abbildung 18 dargestellt sind, geben Hinweise darauf, wo diese Beobachtungen liegen. Von einer kurzen Phase 1987 abgesehen sind Beobachtungen mit hoher geglätteter Wahrscheinlichkeit für Zustand 1 ausschließlich zwischen Anfang 1992 und Ende 1996 zu finden, mehrfach unterbrochen von Zeitpunkten mit niedriger geglätteter Wahrscheinlichkeit. Der Anfang dieses Zeitraum fällt zusammen mit dem Beginn der Krise des EWS. Der damit verbundene Vertrauensverlust des EWS scheint den Mechanismus der PPP also wenigstens teilweise (wieder) in Kraft gesetzt zu haben.

Diese Entwicklung überdauerte die Erweiterung der Bandbreiten im September 1993, die die (schwache) Bindung des Wechselkurses an die Preisentwicklung bestätigte, und endete zu einem Zeitpunkt, zu dem das Vertrauen der Märkte an die Leitkurse wiederkehrte und sich ein Übergang der Leitkurse in die Euro-Relationen nach Beginn der Währungsunion abzeichnete. Die beiden Zustände mit und ohne Einfluß der PPP lassen sich also auch als Phasen niedriger und hoher Glaubwürdigkeit des EWS deuten.

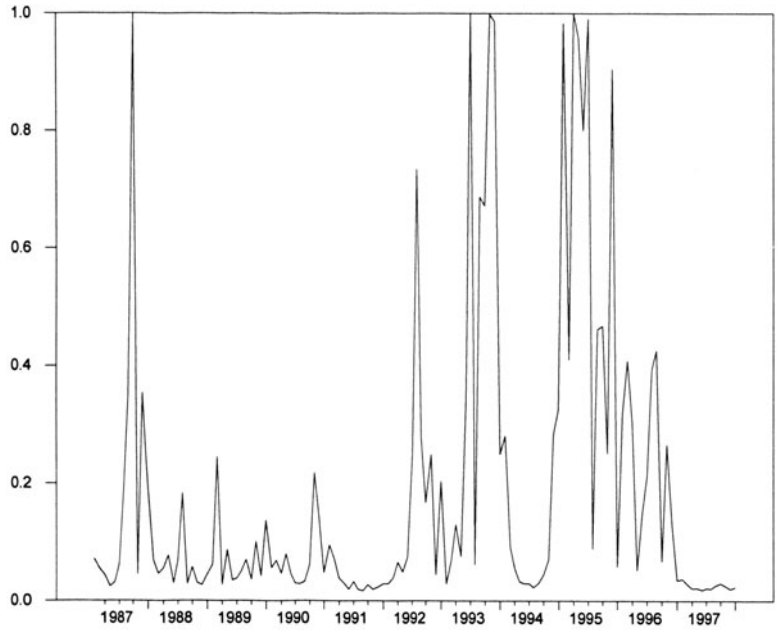

Abbildung 18: Geglättete Wahrscheinlichkeit für Zustand 1 (Markov-Modell)

Es läßt sich daraus jedoch nicht folgern, daß die PPP in dem erwähnten Zeitraum streng Gültigkeit gehabt hätte: Immerhin ist die Standardabweichung des Störterms in diesem Zustand etwa 3 mal so hoch wie in dem anderen. Die PPP war also von starken anderen Einflüssen überlagert.

Die Zeitpunkte mit hoher geglätteter Wahrscheinlichkeit stimmen im wesentlichen mit denen des Modells für die Wochendaten (Abbildung 10, S. 80) überein. Es bestand also ein Zusammenhang zwischen der hohen Volatilität nach der Erweiterung der Bandbreiten und der Tatsache, daß die Inflationsraten wieder Einfluß auf die Wechselkurse gewannen.

11 Schlußbemerkungen

Im methodischen Teil dieser Arbeit wurden die Aspekte von Regimewechselmodellen erläutert, die für empirische Untersuchungen von zentraler Bedeutung sind. Besonderes Gewicht wurde dabei auf zwei Punkte gelegt. Zum einen wurde die Form der Zustandsabhängigkeit der Parameter genau untersucht und Schätz- und Testverfahren vorgestellt, die Modelle betreffen, bei denen einige, aber nicht alle Parameter zustandsabhängig sind, oder einige Parameter nur in einem der Zustände auftreten. Zum anderen wurde die Testtheorie für Regimewechselmodelle systematisch untersucht und an verschiedenen Stellen ausgebaut. Das betrifft etwa den Vergleich verschiedener Ansätze für Tests linearer gegen Regimewechselmodelle, von denen gezeigt wurde, daß sie sich übereinstimmend in einfachen Faustregeln für empirische Untersuchungen zusammenfassen lassen, oder die hier erstmals untersuchte Frage nach Tests zwischen Schwellen- und inhomogenen Mischungs- oder Markov-Modellen, die sich auch für einen Test zwischen getrennten Regimewechselmodellen verwenden lassen.

In den empirischen Untersuchungen dieser Arbeit wurden verschiedene Anwendungsmöglichkeiten von Regimewechselmodellen im Zusammenhang mit Wechselkursdaten erläutert. Die Anwendung der im methodischen Teil entwickelten Teststrategie wurde anhand der empirischen Anwendungen demonstriert. Damit ließ sich unter den vielen in Frage kommenden Modellen jeweils eines auswählen, das sparsam in der Verwendung von Parametern ist und dennoch eine hohe Erklärungsgüte für die Daten besitzt.

Die Ergebnisse zeigen übereinstimmend, daß ein Schwellenmechanismus im Sinne von BEKAERT und GRAY [1998], dessen Schwellenindikator durch die Zinsdifferenz dargestellt wird, keine adäquate Beschreibung der Daten liefert. Obwohl das entsprechende Modell für Wochendaten eine gewisse Erklärungsgüte zu besitzen scheint, haben sich die Markov-Modelle in den empirischen Untersuchungen grundsätzlich überlegen gezeigt.

Mit Hilfe der Markov-Modelle gelang es, verschiedene Charakteristika der Daten in einem einheitlichen Rahmen zu modellieren. Dies betrifft zum einen die statistischen und zeitreihentheoretischen Aspekte, denen in der Finanzmarktökonometrie große Aufmerksamkeit gewidmet wird, also insbesondere die starke Leptokurtosis der Renditen, die stochastische Volatilität, die sich in der Häufung von Beobachtungen mit hoher und mit niedriger Volatilität niederschlägt, und die zeitvariable Autokorrelation der Renditen. Zum anderen wurde demonstriert, daß die Gültigkeit makroökonomischer Zusammenhänge – etwa hier der relativen Kaufkraftparitätentheorie –, die sich in Untersuchungen mit klassischen Regressionsmodellen

empirisch nicht nachweisen lassen, im Kontext von Regimewechselmodellen, die Verhaltensänderungen der beteiligten Subjekte berücksichtigen, in neuem Licht erscheinen kann.

Für all diese Aspekte, die sich mit linearen Modellen nicht erfassen lassen, gibt es jeweils Alternativen zu Regimewechselmodellen. Die Stärke von Regimewechselmodellen liegt aber gerade darin, daß sie einen einheitlichen Zugang zu diesen Gesichtspunkten aus der Zeitreihentheorie, der statistischen Analyse von Renditeverteilungen und der Makroökonometrie gestatten. Mit Hilfe dieser Modelle werden verschiedene *stylised facts* durch *ein* Prinzip erklärt, durch periodische Verhaltensänderungen der beteiligten Akteure.

Da es deutliche Hinweise darauf gibt, daß die Niveaus von Wechselkursen nicht stationär sind, wäre es von Interesse, die Untersuchungen etwa auf integrierte Prozesse auszudehnen. Dies würde die Entwicklung eines völlig neuen Methodenapparates erfordern, der die hier behandelten Verfahren mit der Kointegrationsmethodik verbindet. Erste Ansätze in diese Richtung stellen die Überlegungen in HALL, PSARADAKIS und SOLA [1997] und KROLZIG [1997], S. 297 ff., dar.

Vom empirischen Gesichtspunkt stellt sich die Frage, ob sich die dargestellten Ergebnisse über das EWS auch für den seit Januar 1999 bestehenden europäischen Wechselkursmechanismus, den WKM II, nachweisen lassen. Mit Dänemark und Griechenland sind dem WKM II je ein Land mit als stabil geltender bzw. mit abwertungsgefährdeter Währung beigetreten, so daß eine Konstellation vorliegt, die in mancher Hinsicht mit der hier untersuchten Situation vergleichbar ist.

Literaturverzeichnis

ANDREWS, D. W. K. und PLOBERGER, W. [1994]: Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative, *Econometrica* 62, S. 1383-1414

BANERJEE, ANINDYA [1993]: *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, New York (Oxford University Press, Advanced Texts in Econometrics)

BAUM, LEONARD E.; PETRIE, TED; SOULES, GEORGE und WEISS, NORMAN [1970]: A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *The Annals of Mathematical Statistics* 41, S. 164-171

BEHBOODIAN, JAVAD [1970]: On a Mixture of Normal Distributions, *Biometrika* 57, S. 215-217

BEKAERT, GEERT und GRAY, STEPHEN F. [1998]: Target Zones and Exchange Rates: An Empirical Investigation, *Journal of International Economics* 45(1), S. 1-35

BERNDT, E. K.; HALL, B. H.; HALL, R. E. und HAUSMAN, J. A. [1974]: Estimation and Inference in Nonlinear Structural Models, *Annals of Economic and Social Measurement* 3/4, S. 653-665

BERTOLA, GIUSEPPE und CABALLERO, RICARDO J. [1992]: Target Zones and Realignments, *American Economic Review* 82, S. 520-536

BERTOLA, GIUSEPPE und SVENSSON, L. E. O. [1993]: Stochastic Devaluation Risk and the Empirical Fit of Target Zone Models, *Review of Economic Studies* 60, S. 689-712

BOLDIN, MICHAEL D. [1990]: *Business Cycles and Stock Market Volatility: Theory and Evidence of Animal Spirits*, Dissertation, University of Pennsylvania

BOLLERSLEV, T.; CHOU, R. Y. und KRONER, K. F. [1992]: ARCH Modeling in Finance. A Review of the Theory and Empirical Evidence, *Journal of Econometrics* 52, S. 5-59

BOLLERSLEV, T.; ENGLE, ROBERT F. und NELSON, DANIEL B. [1994]: ARCH Models, in: ENGLE, ROBERT F. und MCFADDEN, DANIEL L. (Hrg.): *Handbook of Econometrics* (*Handbooks in Economics* 2) Vol. IV, Amsterdam (Elsevier Science), S. 2961-3040

BOOTHE, PAUL und GLASSMAN, DEBRA [1987]: The Statistical Distribution of Exchange Rates, *Journal of International Economics* 22, S. 297-319

CARRASCO, MARINE [1997]: *Misspecified Structural Change, Threshold and Markov Switching Models*, unveröffentlichtes Arbeitspapier

CECCHETTI, S. G.; LAM, P.-S. und MARK, N. C. [1990]: Mean Reversion in Equilibrium Asset Prices, *American Economic Review* 80, S. 398-418

CHAN, K. S [1988]: Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *Technical Report* No. 243, Department of Statistics, University of Chicago

— [1990]: Testing for Threshold Autoregression, *The Annals of Statistics* 18, S. 1886-1894

— [1991]: Percentage Points of Likelihood Ratio Tests for Threshold Autoregression, *Journal of the Royal Statistical Society (B)* 18, S. 691-696

— und TONG, HOWELL [1990]: On Likelihood Ratio Tests for Threshold Autoregression, *Journal of the Royal Statistical Society (B)* 52, S. 469-476

CHEUNG, YIN WONG und LAI, KON S. [1993]: Long-Run Purchasing Power Parity during the Recent Float, *Journal of International Economics* 34(1-2), S. 181-92

CLEMENTS, MICHAEL P. und KROLZIG, HANS-MARTIN [1998]: A Comparison of the Forecast Performance of Markov-Switching and Threshold Autoregressive Models of US GNP, *The Econometrics Journal* 1, C47-C75

CLEMENTS, MICHAEL P. und SMITH, JEREMY [1997]: The Performance of Alternative Forecasting Methods for SETAR Models, *International Journal of Forecasting* 13, S. 463-475

COCHRANE, D. und ORCUTT, G. H. [1949]: Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms, *Journal of the American Statistical Association* 44, S. 32-61

COSSLETT, STEPHEN R. und LEE, LUNG-FEI [1985]: Serial Correlation in Discrete Variable Models, *Journal of Econometrics* 27, S. 79-97

COX, D. R. [1961]: Tests of Separate Families of Hypotheses, in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley 1961 (University of California Press)

— [1962]: Further Results on Tests of Separate Families of Hypotheses, *Journal of the Royal Statistical Society (Series B)* 38, S. 45-53

DAVIDSON, RUSSEL und MACKINNON, JAMES G. [1981]: Several Tests for Model Specification in the Presence of Alternative Hypotheses, *Econometrica* 49(3), S. 781-793

DAVIES, ROBERT B. [1977]: Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative, *Biometrika* 64, S. 247-254

— [1987]: Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative, *Biometrika* 74, S. 33-43

DEMPSTER, A. P.; LAIRD, N. und RUBIN, D. [1977]: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society (B)* 39, S. 1-38

DICKEY, DAVID A. und FULLER, WAYNE A. [1981]: Likelihood Ratio Statistics For Autoregressive Time Series with a Unit Root, *Econometrica* 4, S. 1057-1072

DIEBOLD, FRANCIS X. [1988]: *Empirical Modeling of Exchange Rate Dynamics*, Berlin u.a. (Springer, Lecture Notes in Economics and Mathematical Systems 303)

—; LEE, JOON-HAENG und WEINBACH, GRETCHEN C. [1994]: Regime Switching with Time-Varying Transition Probabilities, in: HARGREAVES, COLIN (Hrg.): *Non-Stationary Time Series Analysis and Cointegration*, Oxford (Oxford University Press), S. 284-302

DORNBUSCH, RÜDIGER [1976]: Expectations and Exchange Rate Dynamics, *Journal of Political Economy* 84(6), S. 1161-1176

DREGER, CHRISTIAN [1996]: *Monetäre Modelle der Wechselkurserklärung*, Wiesbaden (Gabler)

DUEKER, MICHAEL J. [1993]: Hypothesis Testing with Near-Unit Roots: The Case of Long-Run Purchasing-Power Parity, *Federal Reserve Bank of St. Louis Review* 75(4), S. 37-48

DUFOUR, J.-M. [1995]: Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables Dynamic Models and Cointegration, *CRDE Working Paper* No. 3695, Université de Montréal

DURLAND, J. MICHAEL und MCCURDY, THOMAS H. [1994]: Duration Dependent Transitions in a Markov Model of U.S. GNP Growth, *Journal of Business and Economic Statistics* 12(3), S 279-288

EBERLEIN, ERNST und KELLER, ULRICH [1995]: Hyperbolic Distributions in Finance, *Bernoulli* 1, S. 281-299

—; — und PRAUSE, KARSTEN [1998]: New Insights into Smile, Mispricing and Value at Risk: The Hyperbolic Model, *Journal of Business* 71, S. 371-405

ENGEL, CHARLES und HAKKIO, CRAIG S. [1994]: The Distribution of Exchange Rates in the EMS, *International Journal of Finance and Economics* 1, S. 55-67

ENGEL, CHARLES und HAMILTON, JAMES D. [1990]: Long Swings in the Dollar: Are They in the Data and Do Markets Know It?, *American Economic Review* 80, S. 689-713

ENGLE, ROBERT F. [1984]: Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics, in: GRILICHES, ZWI und INTRILIGATOR, MICHAEL D. (Hrg.): *Handbook of Econometrics* (*Handbook in Economics* 2) Vol. II, Amsterdam (North-Holland), S. 775-826

— und WATSON, MARK W. [1983]: Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models, *Journal of Econometrics* 23(3), S. 385-400

EVERITT, B. S. und HAND, D. J. [1981]: *Finite Mixture Distributions*, London (Chapman and Hall)

FAIR, R. C. und JAFFEE, D. M. [1972]: Methods of Estimation for Markets in Disequilibrium, *Econometrica* 40, S. 497-514

FAMA, E. F. [1965]: The Behavior of Stock-Market Prices, *The Journal of Business* 38, S. 34-105

FERRI, PIERO und GREENBERG, EDWARD [1992]: *Wages, Regime Switching, and Cycles,* Berlin (Springer)

FIELITZ, BRUCE D. und ROZELLE, JAMES P. [1983]: Stable Distributions and the Mixtures of Distributions Hypotheses for Common Stock Returns, *Journal of the American Statistical Association* 78, S. 28-36

FILARDO, ANDREWS J. [1994]: Business Cycle Phases and their Transitional Dynamics, *Journal of Business and Economics Statistics* 3, S. 299-308

FRENKEL, J. A. [1976]: A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence, *Scandinavian Journal of Economics* 78, S. 200-224

FRIEDMANN, R. [1994]: *Untersuchung der Zinsentwicklung auf Regimewechsel*, Universität des Saarlands, unveröffentlichtes Arbeitspapier

GARCIA, RENÉ [1998]: Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models, *International Economic Review* 39(3), S. 763-788

— und PERRON, PIERRE [1996]: An Analysis of the Real Interest Rate under Regime Shifts, *The Review of Economics and Statistics* 78, S. 111-125

GHYSELS, ERIC [1993]: On Seasonality and Switching-Regime Models, in: *Proceedings of the Business and Economic Statistics Section*, Washington

GIAVAZZI, FRANCESCO und GIOVANNINI, ALBERTO [1989]: *Limiting Exchange Rate Flexibility*, Cambridge/Mass. (The MIT Press)

GIAVAZZI, FRANCESCO; MICOSSI, STEFANO und MILLER, MARCUS [1988]: *The European Monetary System*, Cambridge (Cambridge University Press)

GOLDFELD, STEPHEN M. und QUANDT, RICHARD E. [1972]: *Nonlinear Methods in Econometrics*, Amsterdam (North-Holland, Contributions to Economic Analysis 77)

— und — [1973]: A Markov Model for Switching Regressions, *Journal of Econometrics* 1, S. 3-16

— und — [1976]: Switching Regression Models, in: GOLDFELD, STEPHEN M. und QUANDT, RICHARD E. (Hrg.): *Studies in Nonlinear Estimation*, Cambridge/Mass. (Ballinger), S. 1-96

GOURIEROUX, CHRISTIAN und TROGNON, ALAIN [1983]: Testing Nested or Nonnested Hypotheses, *Journal of Econometrics* 21, S. 83-115

GOURIEROUX, CHRISTIAN und MONFORT, ALAIN [1995a]: *Statistics and Econometric Models,* Vol. 1, Cambridge (Cambridge University Press)

— und — [1995b]: *Statistics and Econometric Models,* Vol. 2, Cambridge (Cambridge University Press)

GRANGER, CLIVE W. J. und TERÄSVIRTA, TIMO [1993]: *Modelling Nonlinear Economic Relationships*, Oxford (Oxford University Press, Advanced Texts in Econometrics)

GRAY, STEPHEN F. [1996]: Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process, *Journal of Financial Economics* 42, S. 27-62

GREEN, WILLIAM H. [1997]: *Econometric Analysis* (3. Aufl.), New Jersey (Prentice-Hall)

HALL, A. D. [1996]: The Mixture of Distributions Hypotheses: The Price Volume Relationship in Australian Futures Markets, *Research in Finance, Supplement* 2, S. 261-280

HALL, STEPHEN G.; PSARADAKIS, ZACHARIAS und SOLA, MARTIN [1997]: Cointegration and Changes in Regime: The Japanese Consumption Function, *Journal of Applied Econometrics* 12, S. 151-168

HAMILTON, JAMES D. [1988]: Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control* 12, S. 385-423

— [1989]: A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* 57, S. 357-84

— [1990]: Analysis of Time Series Subject to Changes in Regime, *Journal of Econometrics* 45, S. 39-70

— [1991]: A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions, *Journal of Business and Economic Statistics* 9, S. 27-39

— [1993]: Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime, in: MADDALA, C. S.; RAO, C. R. und VINOD, H. D. (Hrg.): *Econometrics* (*Handbook of Statistics* Vol. 11), Amsterdam (North-Holland), S. 231-260

— [1994a]: State-Space Models, in: ENGLE, ROBERT und MCFADDEN, DANIEL (Hrg.): *Handbook of Econometrics* (*Handbooks in Economics* 2) Vol. IV, Amsterdam (North-Holland), S. 3041-3080

— [1994b]: *Time Series Analysis*, Princeton (Princeton University Press)

— [1996]: Specification Testing in Markov-Switching Time-Series Models, *Journal of Econometrics* 70, S. 127-157

HANSEN, BRUCE E. [1992]: The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov-Switching Model of GNP, *Journal of Applied Econometrics* 7 (Supplement), S. 61-82

— [1996a]: Erratum: The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov-Switching Model of GNP, *Journal of Applied Econometrics* 11, S. 195-198

— [1996b]: Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis, *Econometrica* 64, S. 413-430

HARVEY, ANDREW C. [1993]: *Time Series Models* (2. Aufl.), New York (Harvester Wheatsheaf)

HILDRETH, C. und HOUCK, J. P. [1968]: Some Estimators for a Linear Model with Random Coefficients, *Journal of the American Statistical Association* 63, S. 584-595

JARCHOW, HANS-JOACHIM [1994]: *Monetäre Außenwirtschaft*, I. *Monetäre Außenwirtschaftstheorie* (4. Aufl.), Göttingen (UTB)

JARROW, R. A. und RUDD, A. [1983]: *Option Pricing*, Homewood (Irwin)

JARQUE, CARLOS M. und BERA, ANIL [1982]: Model Specification Tests: A Simultaneous Approach, *Journal of Econometrics* 20(1), S. 59-82

DE JONG, F. [1994]: A Univariate Analysis of EMS Exchange Rates Using a Target Zone Model, *Journal of Applied Econometrics* 9, S. 31-45

JUDGE, GEORGE G.; GRIFFITHS, W. E.; HILL, CARTER; LÜTKEPOHL, HELMUT und LEE, TSOUNG-CHAO [1985]: *The Theory and Practice of Econometrics* (2. Aufl.), New York (Wiley)

KAEHLER, JÜRGEN und MARNET, VOLKER [1994]: Markov-Switching Models for Exchange-Rate Dynamics and the Pricing of Foreign-Currency Options, in: KAEHLER, JÜRGEN und KUGLER, PETER (Hrg.): *Econometric Analysis of Financial Markets*, Heidelberg, S. 203-230

KARLIN, SAMUEL und TAYLOR, HOWARD M. [1975]: *A First Course in Stochastic Processes* (2. Aufl.), New York (Academic Press)

KIEFER, NICHOLAS M. [1978]: Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model, *Econometrica* 46, S. 427-434

— [1980]: A Note on Switching Regression and Logistic Discrimination, *Econometrica* 48, S. 1065-1069

KIM, CHANG-JIN [1994]: Dynamic Linear Models with Markov-Switching, *Journal of Econometrics* 60, S. 1-22

KON, STANLEY J. [1984]: Models of Stock Returns – A Comparison, *The Journal of Finance* 39 (1), S. 147-165

KROLZIG, HANS-MARTIN [1997]: *Markov Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*, Berlin/Heidelberg (Springer)

KRUGMANN, PAUL [1991]: Target Zones and Exchange Rate Dynamics, *Quarterly Journal of Economics* 106, S. 669-682

LAHIRI, K. und WANG, J.G. [1994]: Predicting Cyclical Turning Points with Leading Index in a Markov Switching Model, *Journal of Forecasting* 13, S. 243-263

LEE, JOON-HAENG [1991]: *Nonstationary Markov Switching Models of Exchange Rates: The Pound-Dollar Exchange Rate*, Dissertation, University of Pennsylvania

LEE, LUNG-FEI und CHESHER, ANDREW [1986]: Specification Testing when Score Test Statistics Are Identically Zero, *Journal of Econometrics* 31, S. 121-149

— und PORTER, R. H. [1984]: Switching Regression Models with Imperfect Sample Separation Information – with an Application on Cartel Stability, *Econometrica* 52, S. 391-418

LEROUX, BRIAN G. [1992]: Maximum-Likelihood Estimation for Hidden Markov Models, *Stochastic Processes and Their Applications* 40, S. 127-143

LINDGREN, G. [1978]: Markov Regime Models for Mixed Distributions and Switching Regressions, *Scandinavian Journal of Statistics* 5, S. 81-91

LÜDEKE, DIETRICH [1973]: Strukturveränderungen des Investitionsverhaltens im Konjunkturverlauf, *Jahrbücher für Nationalökonomie und Statistik* 188, S. 288-319

—; HUMMEL, WOLFGANG und RÜDEL, THOMAS [1989]: *Das Freiburger Modell. Ein mittelfristiges ökonometrisches Vierteljahresmodell für die Bundesrepublik Deutschland mit Simulationen zur Arbeitszeitverkürzung und zur Steuerreform*, Berlin (Springer, Studies in Contemporary Economics.)

LUUKKONEN, RITVA; SAIKKONEN, PENTTI und TERÄSVIRTA, TIMO [1988]: Testing Linearity Against Smooth Transition Autoregressive Models, *Biometrika* 75, S. 491-499

LYE, JENNY N.; MARTIN, VANCE und TEO, LESLIE E. [1998]: Parametric Distributional Flexibility and Conditional Variance Models with an Application to Hourly Exchange Rates, *International Monetary Fund Working Paper Series* WP/98/29

MAGNUS, JAN R. [1978]: Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix, *Journal of Econometrics* 7, S. 281-312

MANDELBROT, B. [1963]: The Variation of Certain Speculative Prices, *Journal of Business* 36, S. 394-419

MARK, NELSON C. [1990]: Real and Nominal Exchange Rates in the Long Run: An Empirical Investigation, *Journal of International Economics* 28(1/2), S. 115-36

MCALEER, MICHAEL [1995]: The Significance of Testing Empirical Nonnested Models, *Journal of Econometrics* 67, S. 149-171

MCLEOD, A. I. und LI, W. K. [1983]: Diagnostic Checking ARIMA Time Series Models Using Squared-Residual Autocorrelations, *Journal of Time Series Analysis* 4, S. 269-273

MITTNIK, STEFAN und RACHEV, SVETLOZAR T. [1993]: Modeling Asset Returns with Alternative Stable Distributions, *Econometric Reviews* 12(3), S. 261-330

MIZON, GRAHAM E. und RICHARD, JEAN-FRANCOIS [1986]: The Encompassing Principle and its Application to Testing Nonnested Hypotheses, *Econometrica* 54, S. 657-677

OBERHOFER, W. und KMENTA, J. [1974]: A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models, *Econometrica* 42, S. 579-590

PAN, MING-SHIUN; CHAN, KAM C. und FOK, CHI-WING [1995]: The Distribution of Currency Futures Price Changes: A Two-Piece Mixture of Normals Approach, *International Review of Economics and Finance* 4(1), S. 69-78

PEARSON, K. [1894]: Contributions to the Mathematical Theory of Evolution, *Philosophical Transactions of the Royal Society* 185, S. 71-110

PFAFF, BERNHARD [1998]: *Stabilitätsüberprüfung von Geldnachfragefunktionen ausgewählter EU-Staaten: eine Darstellung und Anwendung der Flexible Kleinste Quadrate-Methode*, Baden-Baden (Nomos)

PHILLIPS, KERK L. [1991]: A Two-Country Model of Stochastic Output with Changes in Regime, *Journal of International Economics* 31, S. 121-142

POTTER, SIMON M. [1995]: A Nonlinear Approach to US GNP, *Journal of Applied Econometrics* 10, S. 109-125

PRAETZ, P. D. [1972]: The Distribution of Share Price Changes, *Journal of Business* 45, S. 49-55

RAJ, B. und ULLAH, A. [1981]: *Econometrics – A Varying Coefficients Approach*, New York

RITCHEY, ROBERT J. [1990]: Call Option Valuation for Discrete Normal Mixtures, *Journal of Financial Research* 13(4), S. 285-296

ROSE, KLAUS und SAUERNHEIMER, KARLHANS [1995]: *Theorie der Außenwirtschaft* (12. Aufl.), München (Vahlen)

RÜDEL, THOMAS [1989]: *Kointegration und Fehlerkorrekturmodelle. Mit einer empirischen Untersuchung zur Geldnachfrage in der Bundesrepublik Deutschland*, Heidelberg (Physica-Verlag, Wirtschaftswissenschaftliche Beiträge 15)

RUUD, PAUL A. [1991]: Extension of Estimation Methods Using the EM-Algorithm, *Journal of Econometrics* 49, S. 305-341

SCLOVE, STANLEY L. [1983]: Time Series Segmentation: A Model and a Method, *Information Sciences* 29, S. 7-25

SICHEL, D. E. [1994]: Inventories and the Three Phases of the Business Cycle, *Journal of Business and Economic Statistics* 12, S. 269-277

SMITH, G. und SPENCER, M. [1991]: Estimation and Testing in Models of Exchange Rate Target Zones and Process Switching, in: KRUGMANN, PAUL und MILLER, M. (Hrg.): *Exchange Rate Targets and Currency Bands*, Cambridge (Cambridge University Press)

SNEDECOR, GEORGE W. und COCHRANE, WILLIAM G. [1980]: *Statistical Methods* (7. Aufl.), Ames (Iowa State Univ. Press)

TERÄSVIRTA, TIMO [1994]: Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, *Journal of the American Statistical Association* 89, S. 208-218

—; TJØSTHEIM, DAG und GRANGER, CLIVE W. J. [1994]: Aspects of Modelling Nonlinear Time Series, in: ENGLE, R. F. und MACFADDEN, D. L. (Hrg.): *Handbook of Econometrics* (*Handbooks in Economics* 2) Vol. IV, Amsterdam (Elsevier Science)

TITTERINGTON, D. M.; SMITH, A. F. M. und MAKOV, U. E. [1985]: *Statistical Analysis of Finite Mixture Distributions*, New York (Wiley)

TONG, HOWELL [1983]: *Threshold Models in Non-Linear Time Series Analysis*, New York (Springer Lecture Notes in Statistics 21)

— [1990]: *Non-Linear Time Series: A Dynamical System Approach*, Oxford (Oxford University Press)

TYSSEDAHL, J. S. und TJØSTHEIM, D. [1988]: Autoregressive Models With Suddenly Changing Parameters and an Application to Stock Market Prices, *Applied Statistics* 37, S. 353-69

Personen- und Sachregister

Andrews 94
ARCH-Modelle 10, 34, 71, 72, 93, 129
Banerjee 104
Baum 52
Behboodian 54
Bekaert 32, 33, 73, 137
Bera 71
Berndt 82
Bertola 31
bilaterale Leitkurse 22, 23, 25
Boldin 100
Bollerslev 10, 72
Boothe 9, 33, 98, 104
Caballero 31
Carrasco 96, 97, 109
Cecchetti 10
Chan 9, 90, 91, 97, 98
Chesher 95
Cheung 34
Clements 109, 119
Cochrane 44, 71
Cochrane-Orcutt-Verfahren 44
Cosslett 10, 50, 103
Cox 107
Davidson 107
Davies 94, 98, 100
Davies' quick rule 100
Dempster 53, 54, 61
diagnostische Tests 93
Dickey 104
Diebold 14, 28, 34, 48, 55, 57, 63, 64, 65, 66, 67
Dornbusch 30
Dreger 28
Dueker 34
Dufour 95
Durland 14
Eberlein 9
ECU 22, 23, 25
ECU-Leitkurse 22, 32
EM-Algorithmus 50, 51, 52, 53, 54, 55, 56, 57, 58, 61, 63, 65, 81, 82, 112, 120
Engel 10, 33
Engle 10, 33, 53, 57, 72, 93, 94
E-Schritt 53, 55, 56, 57, 65, 81, 82
Everitt 8
Fair 7
Fama 8
Ferri 10
Fielitz 9
Filardo 14, 103
Fok 9
Frenkel 30
Friedmann 11
Fuller 104
Garcia 2, 5, 96, 97, 99, 101, 111
geglättete Wahrscheinlichkeiten ... 63, 77, 122, 125
getrennte Hypothesen 107, 108, 109, 114
Ghysels 14
Giavazzi 21, 23, 24
Giovannini 21, 23, 24
Glassman 9, 33, 98, 104
Goldfeld 6, 7, 10, 11, 51, 99
Gourieroux ... 52, 53, 86, 90, 93, 107, 108, 109
Gradientenverfahren 52
Granger 7, 96, 104
Gray 32, 33, 50, 73, 82, 137
Greenberg 10
Grenander-Bedingungen 85
Hakkio 33
Hall 9, 138

Hamilton....10, 17, 18, 33, 39, 47, 51, 52, 53, 54, 55, 56, 57, 64, 93, 103, 107
Hand 8
Hansen 97, 99
Harvey 90, 108
Hildreth 16
Houck 16
Hummel 4, 16, 17
hyperbolische Verteilung 9
i.i.d.-switching models 8
intramarginale Interventionen.....23, 26, 133
Jaffee 7
Jarchow 27, 29
Jarque 71
Jarrow 70
Judge..... 15, 39, 51, 85, 89, 90, 106, 108
Kaehler 9, 10, 33, 76, 127
Kalman-Filter 18, 57
Karlin 18
Kaufkraftparitätentheorie.....24, 27, 133, 134, 137
Keller 9
Kiefer 54, 92
Kim 57
Kmenta 44, 46
Kointegration 34, 138
Kon 9, 98
Krolzig 17, 61, 109, 120, 138
Kroner 10
Krugmann 27, 31
Krugmann-Modelle 27, 31
Kurtosis 8, 70, 71, 76, 129
Lahiri 10
Lai 34
Laird 53, 54, 61
Lee.....10, 14, 33, 48, 50, 55, 57, 63, 64, 65, 66, 67, 92, 95, 103
Leptokurtosis...8, 9, 72, 75, 76, 127, 130, 137
Leroux 92
Li 71
Likelihood-Quotienten-Tes.....*t*94, 95, 97, 104, 106, 110, 113, 131
Lindgren 54
Lüdeke 4, 16, 17, 105
Luukkonen 98
Lye 9
Magnus 44, 46, 88
Makov 8
Mandelbrot 8
marginale Interventionen.....23, 26, 131
Mark 10, 34
Marnet 9, 10, 33, 76, 127
Martin 9
McAleer 107
McCurdy 14
McKinnon 107, 109
McLeod 71
Micossi 21
Miller 21
Mittnik 9
Mizon 109
ML-Schätzung.....35, 40, 44, 59, 69, 101, 118
Monetäre Wechselkursmodelle 30
Monfort 52, 53, 86, 90, 93, 107, 108, 109
M-Schritt..53, 56, 58, 59, 61, 64, 81, 112
Nelson 10, 72
nuisance parameter 96
numerische Verfahren..... 51, 52, 82
Oberhofer 44, 46, 61
Oberhofer-Kmenta-Verfahren 44, 61
Orcutt 44
Pan 9
Paretoverteilung 9
Paritätengitter..... 22, 23, 26, 32

Pearson 8
Perron 5
Pfaff 15
Phillips 5
Platykurtosis 8
Ploberger 94
Porter 10
Portmanteau-Tests 96
Potter 104, 107
Praetz 9
Prause 9
Psaradakis 138
Quandt 6, 7, 10, 11, 51, 99, 109
Rachev 9
Raj 16
Richard 109
Ritchey 127
Rose 27
Rozelle 9
Rubin 53, 54, 61
Rudd 70
Rüdel 4, 16, 17, 104
Ruud 53, 61
Saikkonen 98
Sauernheimer 27
scale mixtures 9
Schiefe 8, 9, 70, 71, 76, 129
Sclove 47
SETAR-Modelle 7
Sichel 5
simple switching models 8
Simplex-Algorithmus 52
Smith 8, 32, 119
Snedecor 71
Sola 138
Spencer 32
stabile Verteilungen 9
STAR-Modelle 7
Störparameter 96
Strukturbruch 70, 75, 77, 109
Svensson 31
TAR-Modelle 7
Taylor 18
Teräsvirta 7, 96, 98, 104
Titterington 8
Tjøstheim 7, 47, 96
Tong 2, 7, 90, 91, 96, 97, 119
Trognon 107
Tyssedahl 47
überschießende Wechselkursreaktionen 30
Ullah 16
variable Parameter 16
vektorautoregressive Modelle 5
vektorieller Indikator 48, 66
verallgemeinerte Schwellenmodelle 7, 13, 98, 102, 103
Verallgemeinerter EM-Algorithmus 61
Verweildauer 10, 14, 20, 79, 115, 135
volatility clustering 10, 72
Wald-Test 114, 116, 135
Wang 10
Watson 53, 57
Weinbach ... 14, 48, 55, 57, 63, 64, 65, 66, 67
WKM II 25, 26, 138
Zinsparitätentheorie 28, 30